



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research Biological Systems Science Division Update

Todd Anderson, Ph.D.

Director, Biological Systems Science Division, Department of
Energy, Office of Biological & Environmental Research

October 22, 2020

Update on Programmatic Activities

Completed Reviews/Activities

- ✓ Early Career Panel Applications April 30-May 1
- ✓ Quantum Bioimaging Applications June 15
- ✓ Computational Biology FOA Applications June 16-17
- ✓ Sustainability FOA Applications June 18-19
- ✓ National Microbiome Data Collaborative (NMDC) Application August 3
- ✓ Secure Biosystems Design Applications August 10

Upcoming Reviews/Activities

- KBase Review February 3-4
- PI Meeting End of February
- Microbiome SFAs Spring FY21
- COV July 14-15
- ENIGMA August 3-4
- LLNL SFA August 5
- LANL SFA Summer FY21

} Virtual activities

} Tentatively Scheduled as
in-person reviews

NEW FY 2021 Funding Opportunities (Tentative)*

Genomic Science Program

- *Bioenergy Research Centers (BRCs)*
- *Systems Biology for Bioenergy**
- *Plant Biology Research*
- *Sustainability Research for Bioenergy*
- *Biosystems Design*
 - *Secure Biosystems Design*
- *Environmental Microbiome Science*
- *Computational Biosciences*

Biomolecular Characterization and Imaging Science

- *BER capabilities at Light and Neutron Sources*
- *Cryo-EM capabilities*
- *Bioimaging & Quantum Science**

Facilities & Infrastructure

- *Joint Genome Institute (JGI)*

Funding Opportunity Announcements (FOAs)*

Genomic Science Program

Systems Biology for Bioenergy

Biomolecular Characterization and Imaging Science

Bioimaging Research

New Sustainability Research for Bioenergy Awards

Systems Biology Research to Advance Sustainable Bioenergy Crop Development (DE-FOA-0002214)

7 new projects

Novel approaches integrating field studies with computation to understand the complex interactions among plants, soil, and soil microbes that impact bioenergy crop productivity and stress resistance:

- Molecular and cellular networks controlling nitrogen-fixation in aerial roots of sorghum
- Sorghum-AMF interactions to improve sorghum productivity
- Genomics and biochemical approaches to investigate and enhance nitrogen use efficiency and oil yield in Camelina
- Molecular mechanisms of stress resilience in sorghum and *Brachypodium* for increased plant biomass production under nitrogen limitation
- Characterize diazotrophic endophyte communities in *Populus* for improved productivity and resilience
- Predictive models of plant-microbe-environment interactions in switchgrass for climate adaptation and sustainability



\$68M over 5 yrs

<https://www.energy.gov/articles/department-energy-provide-68-million-bioenergy-crop-research>

New Comp Bio Awards

Computational Tool Development for Integrative Systems Biology Data Analysis (DE-FOA-0002217)

6 new projects

Novel computational tools to develop scalable solutions for omics analysis, data mining, and knowledge extraction from complex datasets.

- Finding the Missing Pieces: Filling Gaps that Impede the Translation of Omics Data into Models
- Discovery of Signaling Small Molecules (e.g., Quorum-Sensing Molecules) from the Microbiome
- Machine-Learning Approaches for Integrating Multi-Omics Data to Expand Microbiome Annotation
- High-Throughput, Accurate Gene Annotation Through Artificial Intelligence and HPC-Enabled Structural Analysis
- Harnessing the Power of Big Omics Data: Novel Statistical Tools to Study the Role of Microbial Communities in Fundamental Biological Processes
- Overcoming Systems Biology Bottlenecks: A Pipeline for Metabolome Data Processing Analyses and Multi-Omics Integration



\$5M total

<https://www.energy.gov/science/articles/doe-awards-5-million-computational-biology-research>

New Lab Funding Secure Biosystems Design SFAs and Projects

4 New Secure Biosystems Design Science Focus Areas (SFA)

From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms
LLNL (PI: Yongqin Jiao)

IMAGINE BioSecurity: Integrative Modeling and Genome-Scale Engineering for Biosystems Security
NREL (PI: Michael Guarnieri)

Persistence Control of Engineered Functions in Complex Soil Microbiomes
PNNL (PI: Robert Egbert)

Secure Ecosystem Engineering and Design
ORNL (PI: Paul Abraham)

2 New Secure Biosystems Design Projects

Rapid Design and Engineering of Smart and Secure Microbiological Systems
ANL (PI: Dionysios Antonopoulos)
2-year pilot funding

Intrinsic Control for Genome and Transcriptome Editing in Communities
SNL (PI: Joseph Schoeniger)
3 years of funding

Existing SFA

Microbial Community Analysis & Functional Evaluation in Soils (m-CAFEs)
LBNL (PI: Trent Northen)
(supplement for EcoPOD related research - \$1M)

\$20M provided in **FY20**

New Lab Funding Quantum Science-enabled Bioimaging Projects

New Projects

A Quantum Enhanced X-ray Microscope
BNL (PI: Sean McSweeney)

The 3DQ Microscope: A Novel System Using Entangled Photons to Generate Volumetric Fluorescence and Scattering Images for Bioenergy Applications
LLNL (PI: Ted Laurence)

Quantum Ghost Imaging of Water Content and Plant Health with Entangled Photon Pairs
LANL (PI: James Werner)

Probing Photoreception with New Quantum-Enabled Imaging
PNNL (PI: James Evans)

\$5M/yr over 4 years FY20-FY23

DOE Systems Biology Knowledgebase

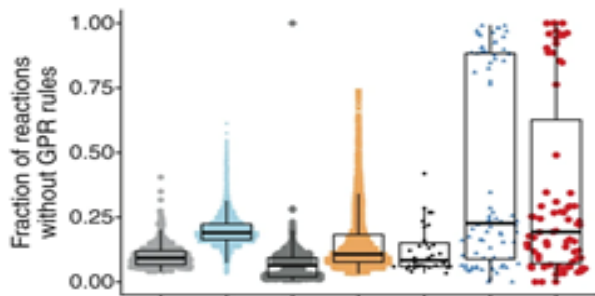
Significant publications and events show the impact of KBase on community research

KBase releases its biochemistry database with support for community chemistry contributions in *Nucleic Acids Research*

Seaver SMD et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* 2020. [doi:10.1093/nar/gkaa746](https://doi.org/10.1093/nar/gkaa746)

Search results for Pyruvate (ID: 34243518). The interface shows a table with columns for ID, Name, Formula, Mass, Charge, and Aliases. A detailed view for Pyruvate is shown below the table, including its chemical structure, formula (C3H4O3), mass (87), charge (-1), and various identifiers like InChIKey, SMILES, and KEGG ID.

In a *Nature Biotechnology* paper about model standards with most community leaders as authors, thousands of models from KBase demonstrated superior diversity and quality scores

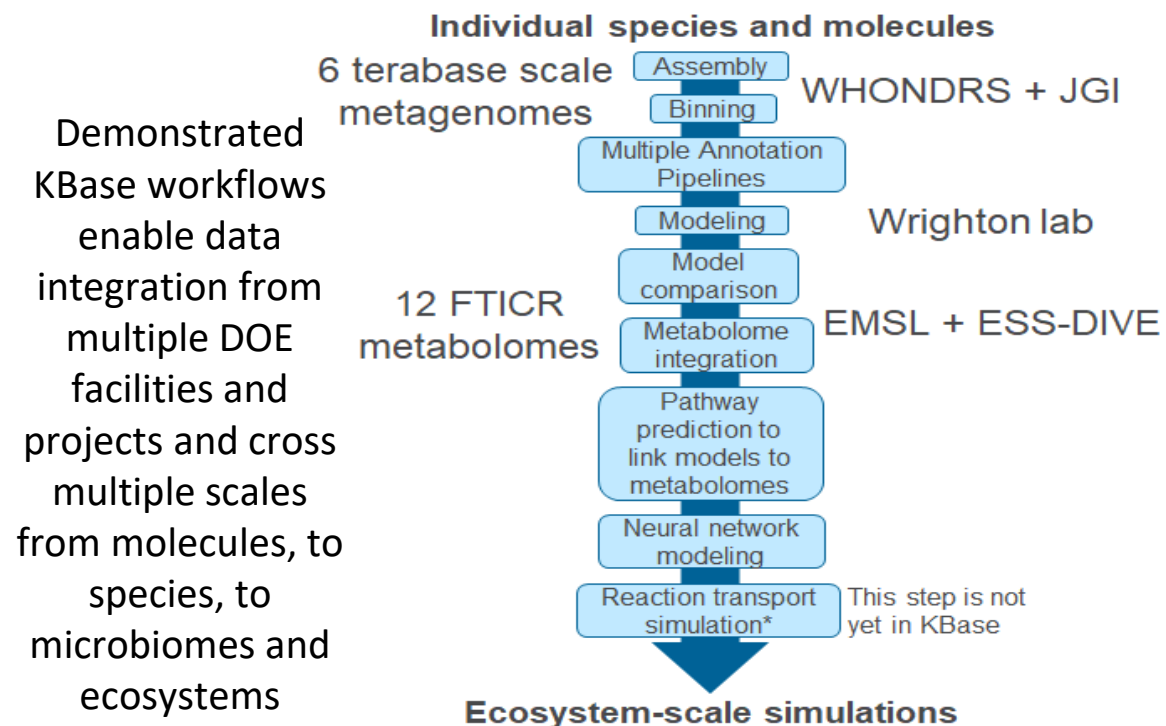


KBase models had fewer reactions without genes



KBase models display diversity

PNNL summer school on multiscale microbial dynamics modeling hosted using KBase



Vision: Empower the research community to harness microbiome data exploration and discovery through a collaborative and integrative data science ecosystem



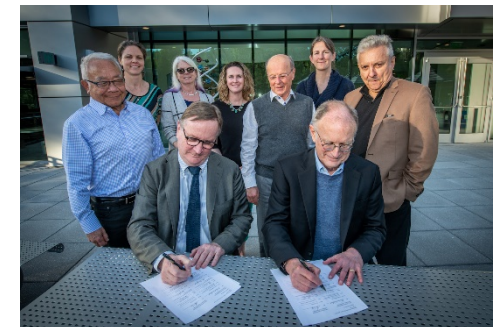
COMMENT

The National Microbiome Data Collaborative: enabling microbiome science

Elisha M. Wood-Charlson¹, Anubhav², Deanna Auberry², Hannah Blanco³, Mark I. Borkum², Yuri E. Corilo², Karen W. Davenport⁴, Shweta Deshpande¹, Ranjeet Devarakonda³, Meghan Drake³, William D. Duncan³, Mark C. Flynn⁴, David Hays¹, Bin Hu⁴, Marcel Huntemann¹, Po-E Li⁴, Mary Lipton², Chien-Chi Lo⁴, David Millard², Kayd Miller³, Paul Piehowski², Samuel Purvine², T.B.K. Reddy¹, Migun Shakra⁴, Jagadish Chandrabose Sundaramurthi³, Pajau Vangay¹, Yaxing Wei³, Bruce E. Wilson³, Shane Canon¹, Patrick S. G. Chain⁴, Kjersten Fagnan¹, Stanton Martin³, Lee Ann McCue², Christopher J. Mungall¹, Nigel J. Mouncey¹, Mary E. Maxon¹ and Emiley A. Eloe-Fadros¹ ¹ES2

Tri-Institutional Partnership in Microbiome Research

Joint \$1M seed funding for five research teams



Accomplishments

- Created the NMDC Data Schema v1.0 and curated 40,000+ environmental biosamples
- Coordinated workflow development with JGI and EMSL for multi-omic data processing
- Initiated community-centered iterative design process for portal development
- Developed Tiered Engagement Strategy and launched the NMDC Champions Program
- Secured funds (\$9M) for additional 12-months

Metadata

<https://github.com/microbiomedata/nmdc-metadata>

Software <https://github.com/microbiomedata/WorkflowPlanning>

<https://hub.docker.com/u/microbiomedata>

Prototype

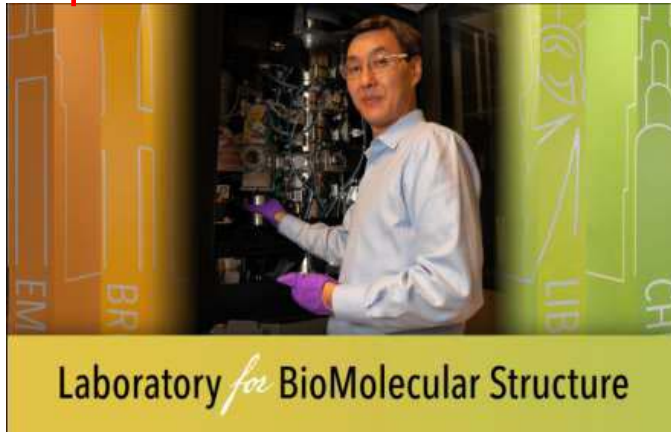
<http://lb.nmdc-dev.development.svc.spin.nersc.org>

Biomolecular Characterization and Imaging Science

Enabling
Capabilities

BNL's cryo-EM Laboratory for BioMolecular Science (LBMS)

- open for business



BROOKHAVEN
NATIONAL LABORATORY



SLAC NATIONAL
ACCELERATOR
LABORATORY



*Eiger 2XE 16M pixel array detector
for SSRL microfocus beamline 12.1,
BSSD co-funded with NIH*

Toward Oilcane: Engineering Hyperaccumulation of TAGs in Sugarcane Stems

Objective

- Divert carbon flux from sucrose into oil in plants to boost biodiesel production. Accumulation of triacylglycerol (TAG) in leaf tissues has been successful in model plants and high biomass crops, but leaf tissues represent only a minor fraction of the biomass of tall C4 grasses. Therefore, accumulation of TAG in stem tissues is highly desirable.

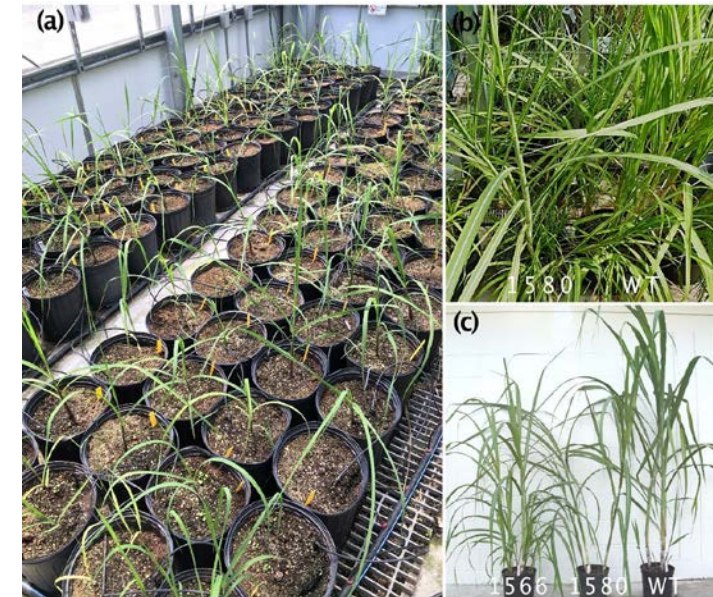
Approach

- Transgenic sugarcane co-expressing WRI1, DGAT1-2 and OLE1 and/or suppressing SDP1 and/or TGD1 was generated.
- Eight transgenic, lipid-accumulating sugarcane lines were compared with non-transgenic sugarcane under controlled greenhouse conditions in a randomized block design with eight replicates.

Results/Impacts

- TAG accumulation in stem tissue of sugarcane was demonstrated, achieving an average of 4.3% of dry weight in a replicated greenhouse experiment.
- The TAG content in leaf tissue was also elevated by more than 400-fold compared to non-engineered sugarcane to reach an average of 8.0% of the dry weight and the amount of total fatty acids (FA) reached 13% of dry weight.

TAG lines growing in replicated greenhouse trial.



(a) Lipid cane lines in randomized blocks.
(b) Highest TAG accumulating oilcane line (1580) and non-transgenic sugarcane (WT).
(c) Transgenic oilcane (1566 and 1580) in comparison to non-transgenic plants (WT; right).

Parajuli, S. et al. 2020. "Towards Oilcane: Engineering Hyperaccumulation of Triacylglycerol into Sugarcane Stems". *GCB Bioenergy*. DOI: 10.1111/gcbb.12684 Zale, J. et al. 2016 "Metabolic Engineering of Sugarcane to Accumulate Energy-Dense Triacylglycerols in Vegetative Biomass." *Plant Biotechnology Journal*, 14, 661–669. DOI: 10:1111

Nitrogen cycling in switchgrass varieties

Objective

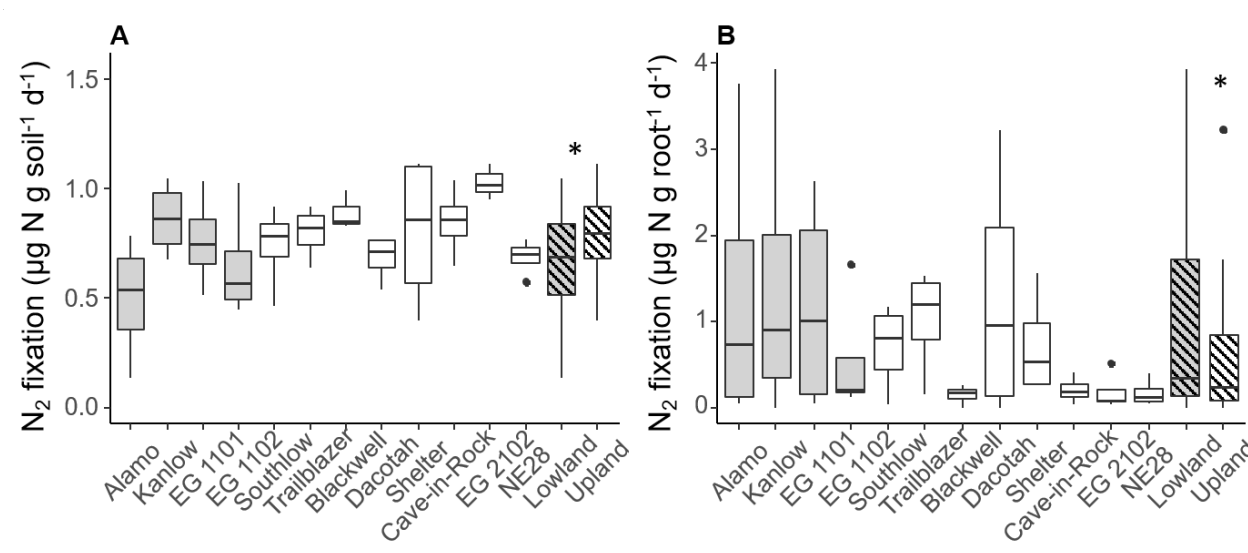
Determine differences in nitrogen (N) acquisition and conservation between switchgrass varieties to optimize growth with the least amount of fertilizer usage.

Approach

Measure N fixation, N mineralization, and N translocation along with soil moisture and N content in switchgrass varieties from upland and lowland ecotypes.

Result/Impacts

- N fixation was highly variable among switchgrass varieties, but in general lowland ecotypes had significantly higher yields, N translocation, and root N fixation, with lower soil N fixation than upland ecotypes.
- N cycling differences might be due to root and soil microbiome differences in switchgrass varieties.
- Many switchgrass varieties can achieve high yield without N additions, decreasing the economic and environmental cost of growing this bioenergy crop.



Novel bacterial clade reveals origin of form I Rubisco

Objectives

- All plant biomass is sourced from the carbon fixing enzyme Rubisco
- Although there is significant interest in engineering Rubisco to improve overall plant productivity, there has been little progress in the field
- All plants contain Form I Rubisco, and all Form I Rubisco consist of both eight Large (RbcL) and eight Small (RbcS) subunits. No Form I Rubisco has previously been discovered to lack the small subunit, which would represent a key transition in Rubisco evolution
- Understanding how Rubisco evolved may provide unique insight into engineering Rubisco

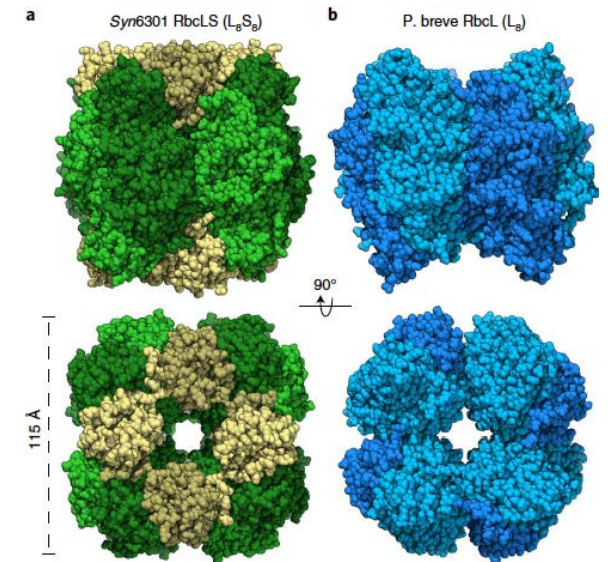
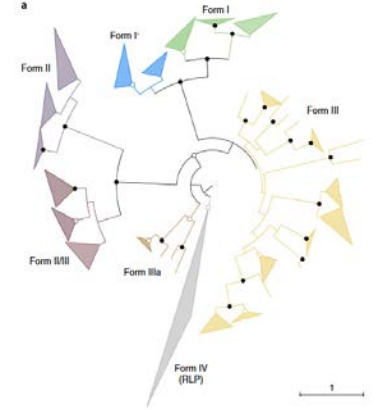
Approach

- Used metagenomics to discover a previously uncharacterized clade of Rubisco that is the sister clade to Form I Rubisco, that we named Form I'
- Synthesized, expressed, purified, and biochemically characterized the novel enzyme
- Solved solution state and crystal structure of novel Form I' Rubisco

Results/Impacts

- Structural insight into how Form I Rubisco evolved will provide the basis for engineering Rubisco in bioenergy crops
- Demonstration of the utility of metagenomics to discover novel families of enzymes

Banda et al., (2020) Novel bacterial clade reveals origin of form I Rubisco, **Nature Plants**, doi.org/10.1038/s41477-020-00762-4



Outer membrane vesicles catabolize lignin-derived aromatic compounds in *Pseudomonas putida*

Objectives

Understand the mechanism of extracellular membrane vesicles in lignin catabolism in bacteria.

Approaches

Evaluate lignin modifications and the exoproteome of three aromatic-degrading bacteria: *Pseudomonas putida* KT2440, *Rhodococcus jostii* RHA1, and *Amycolatopsis* sp. ATCC 39116.

Results/Impacts

- *P. putida* selectively secretes proteins to the extracellular space simultaneous with a reduction in β -O-4 bonds in the lignin substrate.
- Many proteins are encapsulated in outer membrane vesicles, including those involved in the β -ketoacid pathway for aromatic catabolism.
- OMVs from lignin-rich cultivations convert β KA pathway intermediates in vitro and in vivo, demonstrating that enzymes within the OMVs are active.
- We propose a mechanism for lignin-derived aromatic degradation by soil bacteria in the extracellular environment (Fig. B) which holds promise for improving microbial lignin conversion.

Davinia Salvachúa et al., "Outer membrane vesicles catabolize lignin-derived aromatic compounds in *Pseudomonas putida* KT2440," *PNAS* 117:9302-9310; (2020) doi.org/10.1073/pnas.1921073117.

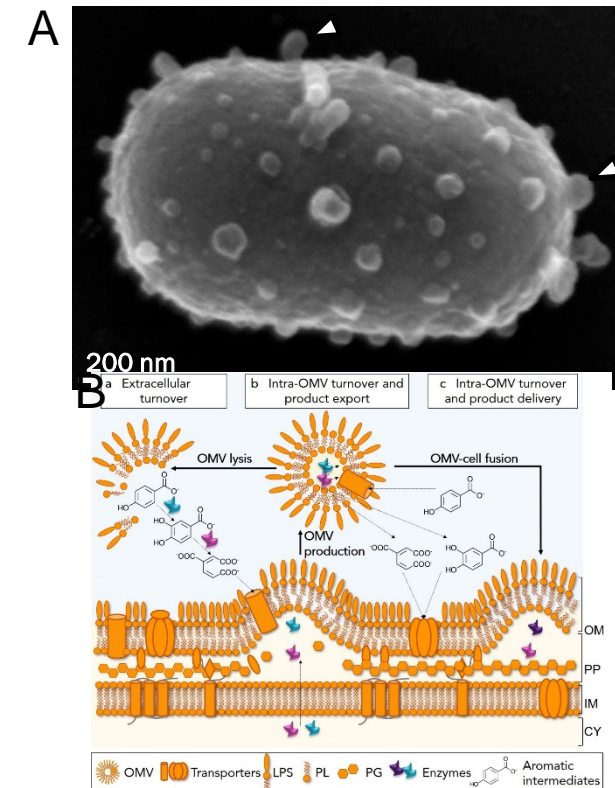


Figure (A) A *P. putida* cell imaged by scanning electron microscopy after 72 h of cultivation in lignin-rich medium where the cell surface is covered with "blebbing" outer membrane vesicles (two blebbing events are highlighted with a white arrowhead). **(B)** Model for OMV-mediated nutrient acquisition and/or catabolism of toxic substrates in *P. putida*.



Biosystems Design Highlight

A Cell-Free Platform to Rapidly Optimize Synthetic Enzymes for Cellular Design

Objective

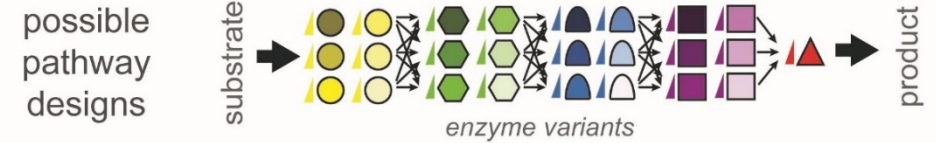
Develop an *in vitro* approach to substantially accelerate the design and optimization of biosynthetic pathways to introduce into engineered cells

Approach

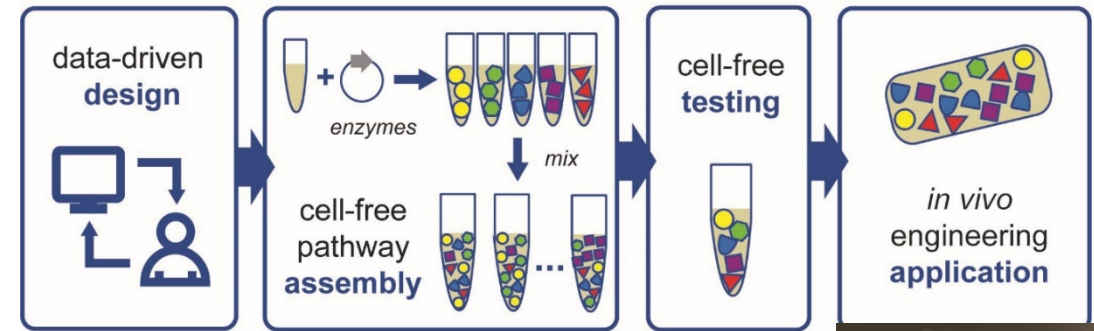
The iPROBE cell-free framework was developed in which cell lysates are enriched with selected enzymes by overexpression or *in vitro* translation. Lysates are mixed to assemble multiple enzymatic pathway combinations. Productivity is then quantified and ranked to select the best pathways to implement *in vivo*

Result/Impacts

- Using iPROBE, 36 pathway combinations were tested for high production of 3-hydroxybutyrate in *Clostridium*, reducing from months to weeks the time needed to develop engineered strains
- Integrated with machine learning iPROBE could rapidly test a large number of enzyme combinations to optimize a six-step n-butanol pathway
- Demonstrated strong correlation between *in vivo* and cell-free pathway performance
- With iPROBE, synthetic pathways can be designed and tested at high throughput, enabling fast engineering of new industrial organisms



The iPROBE Framework



Karim, et al., *Nat Chem Bio* 16, 912–919 (2020)



Differences in substrate use linked to divergent carbon flow during litter decomposition

Objective

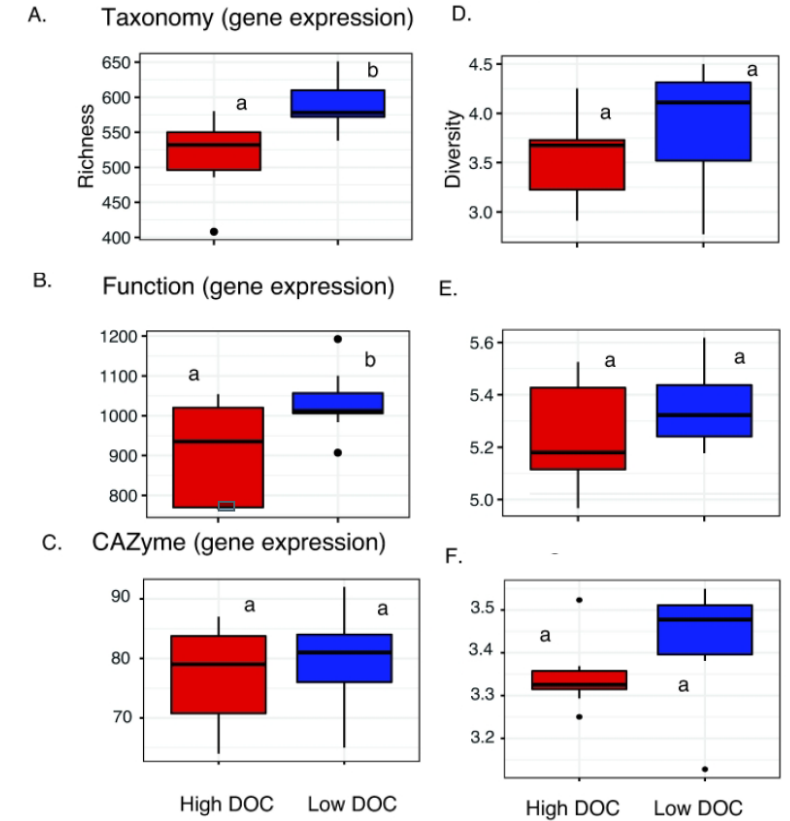
Improve predictions of C flow in terrestrial ecosystems

Approach

- Incubate plant leaf litter with >200 soil microbial communities
- Measure CO₂ and DOC after 44 days of incubation
- Conduct meta-transcriptomics to identify C-cycling metabolic features
- Use machine learning (ML) tools to identify community members and traits that best predict CO₂ / DOC release.

Results/Impacts

- Community composition and taxon richness were strongly correlated with the high DOC release phenotype
- The low DOC phenotype was associated mostly carbohydrate binding modules (CBM) that are used in degradation of complex C such as chitin, xylan, cellulose, and plant lectins, as well as higher fungal signals.
- The high DOC phenotype was associated with expression of pathways for simple sugar breakdown and stress responses.
- Distilling microbial community complexity into a few traits remains an important hurdle to improving model predictions of C flow in terrestrial ecosystems.



Three-dimensional super line-localization in low signal-to-noise microscope images via prior-appraised unsupervised learning (PAUL)

Objective

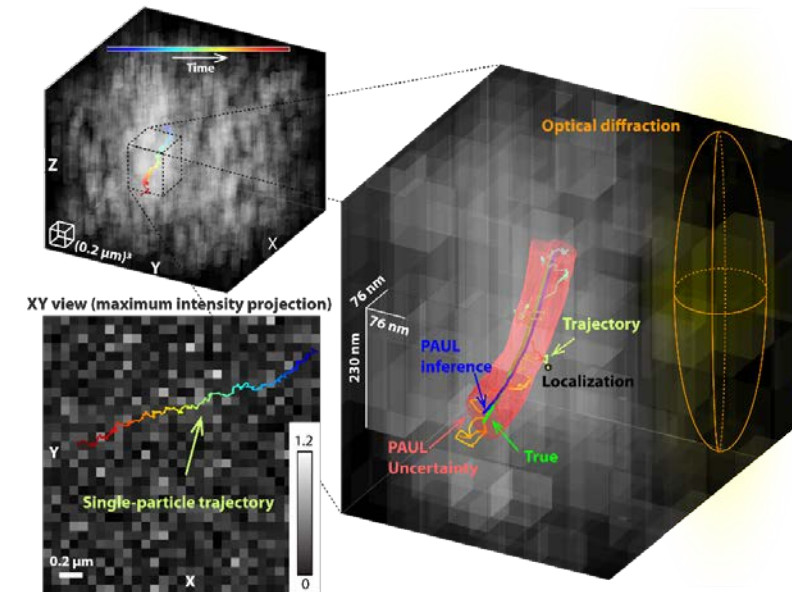
Infer centerline positions of sub-voxel nonparametric curvilinear features from low signal-to-noise 3D images to bridge the gap between single-particle tracking localization and imaging resolution.

Approach

The PAUL framework was applied on 3D images, where each of the three stages was adapted to 3D shapes. The different laterally and axial sampling density has also been considered in stage three where the inter-line angles are calculated in real space.

Results/Impact

- Characterizations using simulated images suggest that by adapting PAUL to 3D line-localization, a medium error of $\sim 1/3$ of the voxel size along z can be achieved.
- This work extracts the feature positions from noisy 3D images that are even difficult to be achieved manually, although human eyes are in general better than machines in recognizing features from noisy pictures.
- The result demonstrates that it is possible to achieve super-line localization in 3D with only the descriptive prior information, similar to 3D super-point localization with the prior information that the signals come exclusively from single emitters.



Shuhui Yin, M. Junaid Amin, Nyssa T. Emerson, and Haw Yang "Three-dimensional super line-localization in low signal-to-noise microscope images via prior-appraised unsupervised learning (PAUL)", *Proc. SPIE* 11510, Applications of Digital Image Processing XLIII, 115101L (21 August 2020); <https://doi.org/10.1117/12.2567752>

Structural Basis for Ca^{2+} -Dependent Activation of a Plant Metacaspase

Objective

Metacaspases are an enzyme family that mediate programmed cell death when plants are stressed. Most metacaspases require Ca^{2+} for activation and action. When Arabidopsis cells are damaged, intracellular Ca^{2+} changes activate Metacaspase 4 (AtMC4), modulating the plant's immune defense. This study sought the mechanism to inform engineering approaches to modulate plant stress response.

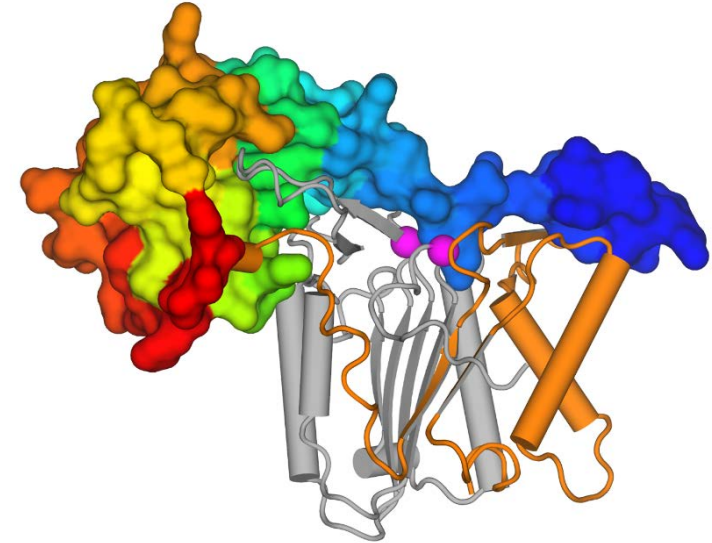
Approach

- AtMC4 structure and its activation process were determined using crystallography
- In vivo activity was analyzed through tobacco plants that were infiltrated with different combinations of the native AtMC4 gene and its mutants, along with substrate

Results/Impacts

Characterized AtMC4 Ca^{2+} -dependent activation:

- In the absence of Ca^{2+} , a linker domain blocks the active site stopping substrate access and processing
- Ca^{2+} elevation causes a conformational change allowing substrate to be processed and initiating a stress response
- Insights gained lay the basis for future engineering by tuning AtMC4 activity to modulate the stress response, enabling biodesign of more sustainable biofuel crops.



Zhu, P., et al. 2020. "Structural basis for Ca^{2+} -dependent activation of a plant metacaspase". *Nature Communications*. 11:2249.
 DOI: 10.1038/s41467-020-15830-8

Targeting Microbial Needles in a Community Haystack

Objective

Enabled by the JGI's Emerging Technologies Opportunity Program, researchers at the Max Planck Institute (MPI) for Marine Microbiology developed, tested and deployed a pipeline to first target cells from communities of uncultivated microbes, and then efficiently retrieve and characterize their genomes.

Approach

Specific bacterial cells are tagged with multiple fluorescent signals, collected through fluorescence activated cell sorting (FACS) and then enriched. These enriched populations are then sequenced and analyzed.

Results/Impact

- Targeted pipeline tested on North Sea metagenomes as a proof of principle in selectively enriching low abundance microbial groups from complex natural communities
- Used a Vis6-specific probe to target bacteria linked to marine diatom blooms. Bacteria which accounted for <1% of the sequences from bulk metagenome reads were amplified to >50% of the target
- Approach could be applied toward gathering more complete view of a population's genetic makeup, and identifying viruses interacting with these populations

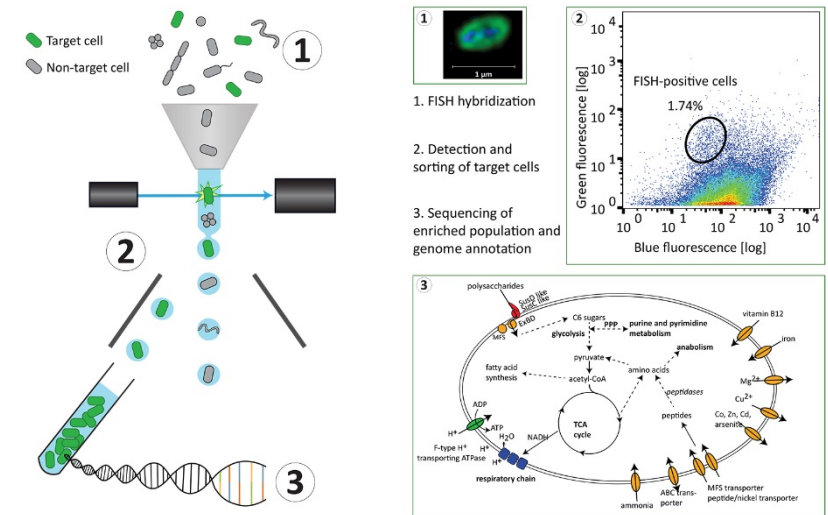


Illustration of protocol developed by Bernhard Fuch's team. (Anissa Grieb; Grieb A et al. *Microbiome*. 2020. CC-BY-4.0)

Grieb A et al. [A pipeline for targeted metagenomics of environmental bacteria](#). *Microbiome*. 2020 Feb 15;8(1):21. doi: 10.1186/s40168-020-0790-7

95 JGI Publications April 1-September 30, 2020

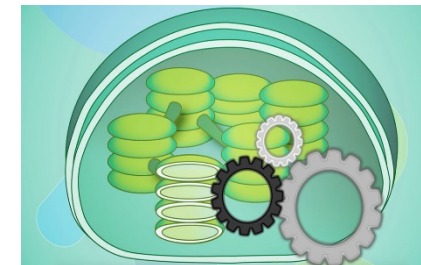
Selected Highlights:

- Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. **ISME J**
- Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. **Nature Communications**
- Bi-directional Titration of Yeast Gene Expression using a Pooled CRISPR Guide RNA Approach. **PNAS**
- Influence of the polar light cycle on seasonal dynamics of an Antarctic lake microbial community. **Microbiome**
- Light-powered CO₂ fixation in a chloroplast mimic with natural and synthetic parts. **Science**
- Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. **Nature Genetics**
- Plant–microbiome interactions: from community assembly to plant health. **Nat Rev Microbiology**

JGI news releases & highlights:

<http://jgi.doe.gov/news-publications/>

<http://jgi.doe.gov/category/science-highlights/>



Thank you

<https://science.osti.gov/ber>

<https://www.energy.gov/science/ber/biological-and-environmental-research>