DOE/SC-0142

Section:
Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities

U.S. DEPARTMENT OF ENERGY
Office of Science

2012

Genomic Science Awardee Meeting X

Bethesda, Maryland
February 26-29, 2012

[Revised: March 1, 2012]

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

http://genomicscience.energy.gov

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

Systems Biology Strategies and Technologies for Understanding Microbes, Plants, and Communities

141

Influence of Habitat on Diversity and Evolution of Surface Ocean Microbes

Andrew E. Allen[1]* (aallen@jcvi.org), Lisa Z. Allen,[1,2] Jonathan H. Badger,[1] Christopher L. Dupont,[1] John P. McCrow,[1] Ian T. Paulsen,[4] Mathangi Thiagarajan,[5] Doug B. Rusch,[5] Jed Fuhrman,[6] Kenneth H. Nealson,[1] Shibu Yooseph, and **J. Craig Venter**[1]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, Calif.; [2]Scripps Institution of Oceanography, University of California-San Diego, La Jolla; [3]Division of Biological Sciences, University of California-San Diego, La Jolla; [4]Macquarie University, Sydney, NSW, Australia; [5]Informatics, J. Craig Venter Institute, Rockville, Md.; and [6]Marine and Environmental Biology, University of Southern California, Los Angeles

Project Goals: The development of computational tools to assess and compare marine metagenomic samples on the basis of phylogenetic and functional diversity, and to examine genome level adaptations to resource availability in different size classes of marine bacteria. The Global Ocean Survey (GOS) continues to sample and analyze marine microbial life around the world from the Sorcerer II research vessel. Comparisons of microbial populations in the Indian Ocean to those found during GOS phase I have resulted in a better understanding of variation in ecological niches available in marine surface waters.

In open ocean pelagic ecosystems bacterial cells account for approximately half of the organic biomass and bacterial activity is responsible for consumption of a large fraction of photosynthetically-derived carbon. Genomic adaptations that underlie bacterial physiological adaptations to environmental gradients such as latitude, temperature, productivity, and particle association have not been clearly elucidated on a large scale. Through analyses of metagenomic data derived from size-fractionated bacterial communities, large scale phylogenomics, and genome size-normalized analyses of genomic contents we elucidate several fundamental adaptations to environmental gradients in the surface ocean. Our results show that marine bacterioplankton in the surface ocean adapt to environmental variability in three ways; molecular-level alterations, changes in overall genomic content, and transitions in community structure. We report on the influence of habitat on genomic properties such as estimated genome size, gene family composition, transporter repertoire, and carbon to nitrogen ratio of the predicted proteome. Small free-living picoplanktonic open-ocean bacterioplankton compared to coastal, larger size-class, or particle associated bacteria have smaller genomes that are enriched for transporters, depleted in regulatory components and encode a proteome with relatively less nitrogen content. Overall patterns of diversity and richness are significantly higher for bacterial communities collected on larger pore-size filters. Intra- and inter-relationships between sites based on diversity measures suggest that larger size class or particle-associated bacterial communities and those that inhabit more productive waters are significantly more variable.

This work was funded by the Office of Biological and Environmental Research in the DOE Office of Science (DE-FC02-02ER63453)

142

Influence of Nutrients and Currents on the Genomic Composition of Microbes across an Upwelling Mosaic

Lisa Zeigler Allen,[1,2] Andrew E. Allen[1]* (aallen@jcvi.org), Eric E. Allen,[2,3] Jonathan H. Badger,[1] John P. McCrow,[1] Ian T. Paulsen,[4] Liam D.H. Elbourne,[4] Mathangi Thiagarajan,[5] Doug B. Rusch,[5] Kenneth H. Nealson,[1] Shannon J. Williamson,[1] and **J. Craig Venter**[1]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, Calif.; [2]Scripps Institution of Oceanography, University of California-San Diego, La Jolla; [3]Division of Biological Sciences, University of California-San Diego, La Jolla; [4]Macquarie University, Sydney, NSW, Australia; and [5]Informatics, J. Craig Venter Institute, Rockville, Md.

Project Goals: Our basic research program focused on acquisition and analysis of marine microbial metagenomic data and development of genomic analysis tools for broad, external community use. Our Marine Metagenomic Diversity effort generated and analyzed shotgun sequencing data from microbial communities sampled from >260 sites around the world. Starting in the summer 2007 and continuing until the winter of 2008, we initiated a series of sampling efforts focused on the Southern California Bight and California Current Ecosystem in collaboration with several groups. In the summer of 2007, this cruise was done in collaboration with the California Cooperative Oceanic Fisheries Investigations (CalCOFI) group to evaluate metagenomics at select areas characterized by coastal upwelling. These samples are a contrast to the primarily open ocean sites of the GOS expedition due to the upwelling of nutrient rich deep waters that enable an increase in primary production at the surface.

Metagenomic datasets were generated from samples collected along a coastal to open ocean transect between Southern California Bight and California Current waters during a seasonal upwelling event, providing an opportunity to examine the impact of episodic pulses of cold nutrient rich water into surface ocean microbial communities. The dataset consists of approximately 5.8 million predicted proteins across seven sites, from 3 different size classes: 0.1-0.8μm, 0.8-3.0μm, and 3.0-200.0μm. Taxonomic and metabolic analyses suggest that sequences from the 0.1-0.8μm size class correlated with their position along the upwelling mosaic. However, taxonomic profiles of bacteria from the larger size classes (0.8-200μm) were less constrained by habitat and characterized by an increase in Cyanobacteria, Bacteroidetes, Flavobacteria, and dsDNA viral sequences. Functional annotation of transmembrane proteins indicate that sites comprised of organisms with small genomes have an enrichment of transporters with substrate specificities for amino acids, iron and cadmium; whereas, organisms with larger genomes have a higher percentage of transporters for ammonium and potassium. Eukaryotic-type glutamine synthetase (GS) II proteins were identified and taxonomically classified as viral, most closely related to the GSII in Mimivirus, suggesting that marine Mimivirus-like particles may have played a role in the transfer of GSII gene functions. Additionally, a Planctomycete bloom was sampled from one upwelling site providing a rare opportunity to assess the genomic composition of a marine Planctomycete population. The significant correlations observed between genomic properties, community structure, and nutrient availability provide insights into habitat-driven dynamics among oligotrophic versus upwelled marine waters adjoining each other spatially.

Publication

1. In Press: Lisa Zeigler Allen, Eric E. Allen, Jonathan H. Badger, John P. McCrow, Ian T. Paulsen, Liam D. H. Elbourne, Mathangi Thiagarajan, Doug B. Rusch, Kenneth H. Nealson, Shannon J. Williamson, J. Craig Venter, Andrew E. Allen. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME Journal.*

This work was funded by the Office of Biological and Environmental Research in the DOE Office of Science (DE-FC02-02ER63453)

143

Genomic Insights to SAR86, an Abundant and Uncultivated Marine Bacterial Lineage

Chris L. Dupont* (cdupont@jcvi.org), Douglas B. Rusch, Shibu Yooseph, Mary-Jane Lombardo, R. Alexander Richter, Ruben Valas, Mark Novotny, Joyclyn Yee-Greenbaum, Jeremy D. Selengut, Dan H. Haft, Aaron L. Halpern, Roger S. Lasken, Kenneth Nealson, Robert Friedman, and **J. Craig Venter**

J. Craig Venter Institute, San Diego Calif.

Project goals: Our basic research program focuses on acquisition and analysis of marine microbial metagenomic data and development of genomic analysis tools for broad, external community use. Our Marine Metagenomic Diversity effort generated and analyzed shotgun sequencing data from microbial communities sampled from >260 sites around the world. Here we used informatic and single cell techniques to generate genomes for uncultivated but abundant organisms, adding genomic context to the metagenomes.

Bacteria in the 16S rRNA clade SAR86 are among the most abundant uncultivated constituents of microbial assemblages in the surface ocean for which little genomic information is currently available. Bioinformatic techniques were used to assemble two nearly complete genomes from marine metagenomes and single-cell sequencing provided two more partial genomes. Recruitment of metagenomic data shows that these SAR86 genomes substantially increase our knowledge of non-photosynthetic bacteria in the surface ocean. Phylogenomic analyses establish SAR86 as a basal and divergent lineage of γ-proteobacteria, and the individual genomes display a temperature-dependent distribution. Modestly sized at 1.25–1.7 Mbp, the SAR86 genomes lack several pathways for amino-acid and vitamin synthesis as well as sulfate reduction, trends commonly observed in other abundant marine microbes. SAR86 appears to be an aerobic chemoheterotroph with the potential for proteorhodopsin-based ATP generation, though the apparent lack of a retinal biosynthesis pathway may require it to scavenge exogenously-derived pigments to utilize proteorhodopsin. The genomes contain an expanded capacity for the degradation of lipids and carbohydrates acquired using a wealth of tonB-dependent outer membrane receptors. Like the abundant planktonic marine bacterial clade SAR11, SAR86 exhibits metabolic streamlining, but also a distinct carbon compound specialization, possibly avoiding competition.

This work was funded by the Office of Biological and Environmental Research in the DOE Office of Science (DE-FC02-02ER63453)

144

Regulation of Nitrogen Metabolism in the Model Marine Diatom *Phaeodactylum tricornutum*

Christopher L. Dupont[1]* (cdupont@jcvi.org), James McCarthy,[1] Adriano Nunes-Nesi,[2] Sabrina Kleeson,[2] Alisdair Fernie,[2] and **Andrew E. Allen**[1]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, Calif. and [2]Max-Planck-Insitut für Molekulare Pfanzenphysiologie, Potsdam-Golm, Germany

Project Goals: While the complete genome sequence of a centric and pennate diatom, forward and reverse genetic techniques and *in silico* modeling have enabled our laboratory and others to begin characterizing unknown

genes, pathways and interactions; nevertheless, key information sets necessary to a systems biology approach to diatom biology remain undeveloped. Our proposed goals focus on two critical gaps in the diatom knowledge base: *i)* although *in silico* models of carbon and nitrogen metabolism depend on information about subcellular locations of metabolic pathway constituents, very few have yet to be experimentally verified; *ii)* overall pathways and mechanisms controlling cellular carbon and nitrogen sensing, assimilation, and flux, in diatoms remain largely undescribed and have not been formally linked to lipid metabolism. Using a combination of transcript and stable isotope metabolite flux profiling in steady state cultures, along with directed enzyme localization and biochemistry experiments, we will evaluate lipid metabolism within the overall context of cellular nitrogen and carbon metabolism.

The unique evolutionary footprint of diatoms may have fostered the evolution of peculiar and unique biochemical pathways contributing to the ecological success of diatoms in the modern ocean. Most notably, a complete metazoan-like urea cycle appears to have been acquired from the host of the secondary endosymbiotic event that gave rise to the Chl c algae. In metazoans, the urea cycle is involved in the catabolism of amino acids and the generation of urea for export. The presence of the urea degrading enzyme urease, acquired from the endosymbiont, strongly suggests an alternative function in diatoms. In marine diatoms, which are frequently subjected to nitrogen limitation, we hypothesize that the urea cycle functions in an anabolic capacity to repackage and recycle inorganic C and N from both endogenous and exogenous sources (Allen et al., 2011). Like green algae and vascular plants, diatom genomes also appear to encode plastid targeted Glutamine Synthetase-Glutamine oxoglutarate aminotransferase (GS-GOGAT) components; unlike green lineage eukaryotes, however, diatoms also express distinct mitochondrially targeted GS-GOGAT genes. This mitochondrial GS-GOGAT cycle, in tandem with a mitochondrial urease, might allow for a rapid redistribution of urea cycle-derived nitrogen metabolites to amino acids following the cessation of nutrient limitation. We propose that a two-part uptake system, involving a plant-like outer membrane transporter and a metazoan-like mitochondrial transporter, delivers urea from the extracellular mileau to the mitochondria. Genomic analyses and metabolite flux studies show that the ammonium produced by urease is assimilated using a complete GS-GOGAT cycle found in the mitochondria, with ancillary fixation through CPS III and the urea cycle. In contrast, nitrate-derived ammonium is clearly assimilated through a plastid-localized GS-GOGAT cycle, with a transfer to the urea cycle metabolite pool via argininosuccinate synthase. Comparative genomic analyses suggests this bifurcated nitrogen assimilation system may be present in other phytoplankton of the chromaveolate lineage. RNAi knock mediated down of mitochondrial and chloroplast localized GS levels are providing additional insights into overall cellular regulation of nitrogen metabolism.

Nitrate reductase (NR) is also enzyme central to overall cellular nitrogen assimilation and metabolism. NR was predominantly believed to be involved in reduction of nitrate as part of nitrogen assimilation. However, mounting evidence suggests a multifunctional role in marine diatoms. First, NR is highly upregulated under cold temperature-high light conditions; this been hypothesized to suggest that NR provides an alternative electron sink for photosynthetically derived electons and reductants that are in excess due to an imbalance between carbon assimilation and growth (Lomas and Gilbert 1999; Parker and Armbrust 2005).. Additionally, in NR-YFP transgenic overexpressors, nitric oxide production is greatly increased. This signaling molecule has been implicated in apotosis and cell-cell signaling in diatoms; although the source of NO in plant cells remains controversial, the peroxisome and NR have each independently, but never together, been implicated in NO production. It is tempting to speculate that NR could be fueling NO production in diatom peroxisomes. In any case, it appears clear that NR is at the center of nitrogen assimilation, signaling, and energy balance. In order to investigate this in more detail we have performed a series of immunolocalization experiments intended to examine NR localization *in vivo* in response to cellular nitrogen status and nitrogen source. RNAseq experiments aimed at preliminary characterization of the diatom transcriptome in response to cellular nitrogen status and nitrogen source have also been performed. In conjunction with the various data types collected to date, an intital genome-scale model of nitrogen metabolism has been constructed.

References

1. Allen, A.E., Dupont,C.L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H., Johnson, D.A., Hu, H., Fernie, A.F., Bowler, C. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473(7346):203-7
2. Lomas, M. W., and P. M. Gilbert. 1999. Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. Limnology and Oceanography 44:556-572.
3. Parker, M. S., and E. V. Armbrust. 2005. Synergistic effects of light, temperature, and nitrogen source on transcription of genes for carbon and nitrogen metabolism in the centric diatom *Thalassiosira pseudonana* (Bacillariophyceae). Journal of Phycology 41:1142-1153.
4. Lomas, M. W., and P. M. Gilbert. 1999. Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. Limnology and Oceanography 44:556-572.

This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science.

145

Responses of and Interactions Between Nitrifying Bacteria to Environmental Changes: A Systems Level Approach

Peter Bottomley* (peter.bottomley@oregonstate.edu), Jeff Chang, Yannming Di, José Perez, Luis Sayavedra-Soto, and **Daniel Arp**

Oregon State University, Corvallis, Ore.

Project Goals: In nitrification, ammonia oxidation and nitrite oxidation are usually coupled; however it is not known to what extent ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB) interact with each other. Two model organisms, *Nitrosomonas europaea* (AOB) and *Nitrobacter winogradskyi* (NOB) are being used to test the hypothesis that AOB and NOB interact with each other, affecting their growth rate, cell yield, and protein and gene expression, and that these effects are distinct from the individually cultured bacteria. Their genome inventories and physiological and gene expression responses to environmental changes, singly and in co-cultures, will be used to construct genome-scale metabolic pathway models. We will create predictive models of AOB and NOB that incorporate metabolism, the regulatory interactions that influence metabolism, and the signaling networks for interaction with the environment. The models will help predict how AOB and NOB will behave in response to changes in the environment.

This project started October 2011. During this time we established the protocols to culture *N. europaea* and *N. winograskyi*, singly and in coculture. The cells are being grown in chemostats and in pH controlled batch fermentors.

A viable chemostat with a coculture of *N. europaea* and *N. winograskyi* was obtained within a month and has been successfully maintained over four months under constant conditions. The hydraulic retention time of 8 days was established and cells were harvested for analysis, allowing for recovery between samplings to minimize variation. From these coculture samples total RNA was isolated and is being used to determine the whole-genome transcript levels. The batch cultures and cocultures were grown in similar medium and at a constant pH. The cells in the batch cultures were harvested at late logarithmic phase. Similarly, total RNA was isolated and is being used to determine the whole-genome gene expression levels.

We are using RNA-Seq to quantify gene expression levels, to detect all messages (including small RNAs) and to avoid cross-hybridization artifacts. An optimized protocol to discriminate the transcriptomes of *N. europaea* and *N. winogradskyi* cocultures is being applied. The computational and statistical methods allow discriminating of the gene expression levels of the two model organisms (Table 1).

Table 1: The large majority of RNA-Seq reads can be uniquely assigned in the genomes of *N. europaea* and *N. winogradskyi*.*

CDS	Possible 25mers	Unique to itself	% to itself	Unique to other	% to other
N. europaea	2,345,948	2,224,679	94.83%	2,224,640	94.83%
N. winogradskyi	2,785,084	2,638,289	94.73%	2,638,250	94.73%

*Unique 25mer sequences from either the concatenated coding sequences (CDSs) per genome were identified. Unique is defined as a sequence found only once (unambiguous). We attempted to align the unique 25mers against the concatenated coding sequences (CDSs) or genome of the other species with no allowance for mismatches. We selected a 25mer for the comparison since most errors in the ~36-bases reads of RNA-Seq are at either end. The reads are trimmed prior to the alignments.

We designed a method to monitor cell viability of *N. europaea* and *N. winogradskyi* while in coculture using allylthiourea (ATU). ATU is a selective inhibitor of ammonia oxidation but not of nitrite oxidation. In this method the ammonia- and nitrite-dependent oxygen uptake are monitored to assess the overall cell health and abundance of each population. The method also allows the monitoring of AOB and NOB activities in coculture through growth, presenting an advantage over monitoring nitrite or nitrate accumulations solely. With this method the relative cell abundances of *N. europaea* and *N. winogradskyi* during coculture growth can be determined within 10% accuracy.

We anticipate that analysis of transcriptomes, growth rates, and metabolites will provide insights to the responses of cells when in co-cultures. We anticipate that changes in gene expression will reflect how cells are sensing and responding to the environmental changes we impose on the co-cultures. The metabolic models will provide a basis to predict and analyze responses of the cell types in co-culture.

Funding for this work is through the U.S. Department of Energy and through the Experimental Station at Oregon State University.

146

Regulons, Uptake, and Salvage Pathways that Mediate Metabolite Exchange in Hot Spring Microbial Mats

Dmitry A. Rodionov[1]* (rodionov@burnham.org), Frank Collart,[2] Aaron T. Wright,[3] Margie Romine,[3] Jim Fredrickson,[3] and **Andrei L. Osterman**[1]

[1]Sanford-Burnham Medical Research Institute, La Jolla, Calif.; [2]Argonne National Laboratory, Argonne, Ill.; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area (FSFA) is to develop a predictive, genome-enabled understanding of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. As part of this multi-institutional effort, we are utilizing genomic sequence to identify candidate systems that mediate and control the synthesis and secretion of chemical commodities by the autotroph and the cor-

responding systems that enable their uptake and salvage by interacting heterotrophs. By compiling this knowledge and employing various experimental methods to validate novel functional predictions, we intend to improve our ability to predict the nature of beneficial partnerships that occur between microorganisms that occupy the same niche.

Research is being conducted on three separate microbial communities; phototrophic mats that occur in alkaline siliceous hot springs in Yellowstone National Park (YNP) or a meromictic, hypersaline lake in northern Washington state and chemotrophic mats that occur in acidic hyperthermophilic springs in YNP. As sequence becomes available from organisms (or their near relatives) that occupy these communities we use integrative genomics-based reconstruction and experimental assessment of metabolic and regulatory networks to understand metabolic interactions and associated regulatory interactions in the target microbial communities. Our general workflow includes: (i) the subsystems-based genomic reconstruction of selected pathways and regulons; (ii) identification of currently unknown components of these pathways, primarily transporters, sensory and regulatory proteins using comparative genomics and high-throughput *-omics* data; (iii) experimental testing of selected functional predictions combining biochemical and genetic methods; and (iv) use model co-cultures to assess metabolic exchange in microbial communities.

Our current research focuses on members of the Cyanobacteria and Chloroflexi phyla that were isolated from or related to isolates detected in YNP phototrophic mats. In these mats, anoxygenic phototrophic bacteria (e.g., *Roseiflexus*, *Chloroflexus*) are thought to function predominantly as photoheterotrophs using metabolites produced by cyanobacteria (e.g., *Synechococcus*) as carbon sources. Whereas typical metabolic byproducts may provide a major flux of carbon and energy, vitamins (precursors of key cofactors) are required in relatively small amounts. We hypothesize that vitamin exchange may be rather widespread phenomena contributing to "opportunistic" relationships between species in mat communities. In a previous genomic survey, we observed a mosaic distribution of *de novo* and salvage pathways for biogenesis of major vitamins (such as niacin, pantothenate, biotin, thiamin, riboflavin) that leads to the presence in communities of the strict auxotrophs and strict prototrophs with respect to one or another vitamin. In this work, we focused on reconstruction of metabolic pathways and regulons involved in metabolite salvage and exchange in two groups of environmental bacteria, Chloroflexi and Cyanobacteria. Several co-cultures (e.g., *Synechococcus sp.* PCC 7002 and *Roseiflexus castenholzii*) have been established by combining isolates available in pure culture and thus provide tractable models for studying naturally occurring interactions.

In Cyanobacteria, we used the comparative genomic approach to reconstruct regulons for 18 transcription factors (including 14 previously known in the literature) and 10 riboswitches. This approach led to the discovery of a putative B_{12} transport uptake system (BtuFCD) in *Synechococcus sp.* PCC 7002. This strain lacks the ability to synthesize B_{12} and thus is dependent on acquiring it from other microbes that are able to produce it. The transporter prediction will be validated through knock-out construction (in collaboration with Don Bryant) and *in vitro* assay. In addition to using regulon/subsystem analysis to discover additional novel B_{12} transporters we will use chemical probe profiling to discover them experimentally.

In Chloroflexi, a repertoire of all transcription factors and environmental sensors encoded in 5 sequenced genomes from this phylum was identified and analyzed. Several bioinformatics workflows are used for identification of conserved DNA and RNA motifs and *ab initio* reconstruction of associated regulons. Currently this approach resulted in reconstruction of 20 regulons operated by either transcription factors or riboswitches, and the number of reconstructed regulons is growing.

The reconstructed regulons lead to the discovery of multiple novel uptake transport systems for essential vitamins (thiamin/B_1, riboflavin/B_2, B_{12}), metal ions (manganese, zinc, ferrous iron, molybdenum, cobalt), and carbohydrates (e.g., rhamnose). These novel transport systems fill the gaps in the respective metabolic pathways. Several examples of novel vitamin uptake systems were selected for further experimental validation. First, a putative B_2 transporter (RibXY) complements the absence of the *de novo* riboflavin biosynthesis pathway in *Roseiflexus* spp. The riboflavin transfer to *Roseiflexus* could be important in co-culture with B_2-prototrophic strains *Synechococcus* and/or *Thermochlorobacter*. Second, we identified putative B_1 precursor transporters (ThiXYZ, ThiW, CytX) that are required for B_1 salvage pathways in the auxotrophic bacteria from the Chloroflexi phylum. The ABC transporter predictions (ThiY, RibY) will be validated through *in vitro* assay. Third, we found a novel ECF-type transport system of yet unknown specificity that is involved in B_{12} metabolism in Chloroflexi. Finally, we found multiple predicted transport systems that are involved in the uptake of biometals (Fe, Zn, Mn, Co, Mo). They are regulated by committed transcription factors to maintain the cell homeostasis.

L-rhamnose is an essential component of the external polysaccharide layer in the cyanobacterial cells. Analysis of *Chloroflexus/Roseiflexus* genomes revealed a conserved gene cluster encoding enzymes from the rhamnose catabolic pathway. We hypothesize that transfer of carbon from YNP *Synechococcus* spp. to YNP *Chloroflexus* via L-rhamnose is an example of metabolite exchange that occurs in native YNP mats. We have identified a novel ABC-type transport system (RhaGHJF) and a DeoR-type transcription factor (RhaR) encoded in the rhamnose gene cluster and propose that they are involved in uptake and sensing of L-rhamnose to induce the catabolic pathway. The RhaR DNA binding sites were predicted and the respective rhamnose regulon was reconstructed by comparative genomics. Experimental validation of predicted functional roles of the novel rhamnose transporter and regulator in a selected model strain of *Roseiflexus castenholzii* is ongoing. In addition, we intend to test the hypothesis of L-rhamnose exchange between Cya-

nobacteria and Chloroflexi in a model co-culture combining gene expression, proteomics and metabolomics analyses.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

147

Methanotrophic-Mediated Metal Binding: Effects on *In Situ* Microbial Community Structure and the Sustainability of Subsurface Water Systems

Jeremy D. Semrau[1]* (jsemrau@umich.edu), Alan A. DiSpirito,[2] and Nathan Bandow[2]

[1]University of Michigan, Ann Arbor and [2]Iowa State University, Ames

http://sitemaker.umich.edu/methanobactin

Project Goals: Methanotrophs are ubiquitous in the environment, and despite their critical function in many different ecosystems, the biogeochemical factors that affect their activity and community structure are poorly understood. It is known that copper plays a key role in methanotrophic physiology, but the mechanism used by these microbes for copper acquisition was only recently discovered. This compound, methanobactin, is the first example of a "copper-siderophore", or chalkophore. Like siderophores, methanobactin binds many different metals, including mercury. Further, as found with siderophores, recent data show that different methanotrophs make different forms of methanobactin that have varying metal affinities. The general objectives of this proposal are thus to consider how methanobactin made by different methanotrophs alters the bioavailability of metals of concern to the DOE and how this affects: (1) the physiology, metabolism and gene expression in pure cultures of methanotrophs; (2) the broader microbial community structure and meta-transcriptome in laboratory soil columns, and; (3) the bioavailability and risk associated with different metals in subsurface environments.

One of the persistent and substantial problems in remediation of hazardous waste sites is the mobilization and uncontrollable transport of radionuclides and heavy metals from these sites to surrounding areas. Some microbially-mediated processes can at least temporarily immobilize and reduce the toxicity of these materials through dissimilatory reduction that leads to precipitation and sorption under anaerobic conditions. As such, microbial-mediated processes can limit the dispersal of these materials and thus also limit the exposure of surrounding areas. Microorganisms, however, have effective and ubiquitous mechanisms to solubilize different metals and that non-specific binding of metals by these biogenic metal chelators may increase their solubility, mobility, and bioavailability. Here we propose to consider how the expression of metal chelating agents analogous to siderophores in methane-oxidizing bacteria i.e., methanotrophs, alters the bioavailability of various metals (e.g., copper and mercury) and how this: (1) affects the physiology, metabolism and gene expression in methanotrophs; (2) affects the broader microbial community structure and meta-transcriptome, and; (3) increases the bioavailability and risk associated with various metals. Such studies will enable us to determine how methanotrophic activity may affect the structure of subsurface microbial communities as well as the sustainability of subsurface waters, including at DOE sites.

This project, starting in September 2011, has as its immediate objective the characterization of metal binding by methanobactin produced by different methanotrophs. The molecule has been identified in a number of methanotrophs, but has only been structurally characterized from Methylosinus trichosporium OB3B (mb-OB3b) and Methylocystis strain SB2 (mb-SB2). mb-SB2 shows a significant similarity to mb-OB3b, including spectral and metal binding properties, and the ability to bind and reduce Cu(II) to Cu(I). Both forms of methanobactin contains similar five-member rings and associated enethiol groups, which together, form the metal ion binding site. Further, greatest amounts of methanobactin are found when both Methylosinus trichosporium OB3B and Methylocystis strain SB2 are grown in low (<1 μM) copper. Significant differences exist between the two known forms of methanobactin (Figure 1), including: (1) the number and types of amino acids used to complete the molecule, (2) mb-OB3b has two oxazolone rings for copper binding, while mb-SB2 has one imidazolone ring and one oxazolone ring, and; (3) a sulfate group is found in mb-SB2 but not in mb-OB3b. The sulfate in mb-SB2 is bonded to a threonine-like side chain and may represent the first example of this type of sulfate group in a bacterial-derived peptide. The gene sequence for a ribosomally produced precursor for mb-OB3b has also been identified in the genome of *Methylosinus trichosporium* OB3b. It indicates that the oxazolone rings in mb-OB3b are derived from glycine and cysteine residues. Taken together, the results reported here suggest methanobactins are a structurally diverse group of ribosomally-produced, peptide-derived molecules that share a common pair of five-member rings with associated enethiol groups that bind and reduce copper in aqueous environments.

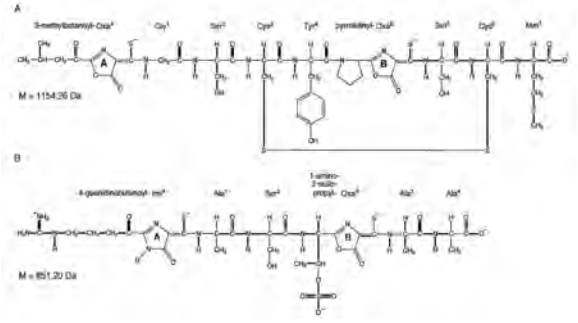


Figure 1. Primary structure of methanobactin from (A) *M. trichosporium* OB3b and (B) *Methylocystis* strain SB2.

* Presenting author

Comparison of the spectral and thermodynamic properties of mb-SB2 demonstrate that the spectral and basic copper binding properties of both methanobactins were similar and the metal binding capacity, particularly for copper, of both methanobactins lies primarily, if not solely, in the pair of five-membered rings and associated enethiol groups. To date, we have surveyed the range of metals bound by both forms of methanobactin, and have found that both forms bind many different metals, including mercury, chromium, lead, and uranium, and that metals are typically reduced after binding by methanobactin.

Next steps will characterize the effect of methanobactin-mediated metal speciation on activity and transcriptome of pure cultures of methanotrophs. In addition, we will collect groundwater and soil samples from the Integrated Demonstration Site of the DOE Savannah River Site where methanotrophs are known to exist and characterize how methanobactin affects metal mobility and bioavailability in the presence of soils from this site as well as methanobactin-mediated dissolution of soil-associated minerals using a set of spectroscopic and imaging techniques. The resultant effects on the broader microbial community structure and function will be also be determined using a "double-RNA" approach to characterize microbial community structure and function simultaneously from the meta-transcriptome.

148

Improved Understanding of Microbial Iron and Sulfate Reduction Through a Combination of Bottom-Up and Top-Down Functional Proteomics Assays

Ruth E. Richardson[1]* (rer26@cornell.edu), Annette R. Rowe,[1] Peter Kelly,[1] Hening Lin,[1] Sheng Zhang,[1] Robert Sherwood,[1] Lee Ann McCue,[2] Michael Wilkins,[2] and Stephen J. Callister[2]

[1]Cornell University, Ithaca, N.Y. and [2]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Our overall project objectives are to improve our understanding of the systems biology of the dissimilatory iron reducing bacteria (DIRB) and sulfate reducing bacteria (SRB) and to use this knowledge to aid in site-monitoring and management via informative biomarkers. The work will employ a wide range of proteomic and biochemical assays. These include both "bottom-up" and "top-down" approaches. In the bottom up approaches, gene targets are selected for heterologous expression based upon proteomic expression patterns and are then put through a range of *in silico*, proteomic and biochemical tests. In the top-down approach, we will develop a high-throughput functional protein assay that screens for function first before then identifying the responsible enzymes– thereby focusing mass spectrometry resources on the identification of functionally-relevant oxidoreductase enzymes and their interactors.

The advent of the genomics era has advanced our understanding not only of the genetic makeup of a wide range of microorganisms (and microbial communities) it has also enabled remarkably high throughput monitoring of gene expression for thousands of genes simultaneously. With current *in silico* bioinformatic tools, inferences can often be made as to the function of some of these gene targets simply from gene sequences or (in the case of "hypothetical" proteins) from expression patterns under different conditions. However, inferences are only the start with respect to learning the true function of poorly annotated gene targets. To turn these observations and inferences into an improved understanding of the systems biology of organisms of interest, functional annotation techniques and approaches must be improved.

We have created an experimental plan to tackle this functional annotation bottleneck in two groups of microorganisms of particular interest for heavy metals and radionuclide bioremediation: the dissimilatory iron reducing bacteria (DIRB) and the sulfate reducing bacteria (SRB). We anticipate that our approaches will both generate a better fundamental understanding of the systems biology of the DIRB and SRB, and lead to improved diagnostic tools for application at contaminated field sites—namely peptide biomarkers that can be used to document and enhance *in situ* activity. To this end we are particularly interested in improved functional characterization of oxidoreductase enzymes as they are key at the biotic/abiotic interfaces in aquifers.

The work is taking take place in three complementary phases. In Phase 1 ("bottom-up"), a combination of *in silico*, proteomic, and biochemical approaches are being applied to phylogenetically-diverse DIRB and SRB cultures (grown under different operating conditions of type of electron acceptor and rate of respiration). This phase will first identify the core proteomes specific to iron and sulfate reduction, then will focus on improved functional characterization of this core proteins via: follow-up *in silico* analysis of sequences, follow-up proteomics to identify protein-protein interaction partners, and biochemical assays focused on documentation of redox capabilities. In Phase 2 ("top-down"), we are developping a high-throughput functional redox protein assay consisting of three steps: non-denaturing separation of a (meta)proteome, parallel assaying of wells for desired redox activity, and tandem mass spectrometry identification of proteins in positive wells. In Phase 3, proteotypic peptides will be elucidated for a suite of functionally-informative enzymes (already known and newly discovered in Phases 1 and 2) and Mass Western assays will be developed for the simultaneous quantification of dozens of such peptide biomarkers in a single nanoLC-MS/MS injection. In Phases 2 and 3, assays will first be tested on lab culture proteomes and then used to quantify biomarkers in biomass samples from sites undergoing in situ bioremediation (e.g. the Rifle site). In our poster presentation we will highlight work performed in the first six months of the project.

This work is fully funded through the DOE Office of Biological and Environmental Research within the Office of Science.

149

In situ Spectroscopy on Intact *Leptospirillum ferrooxidans* Reveals that Reduced Cytochrome 579 is an Intermediate in the Iron Respiratory Chain

Robert C. Blake II* (rblake@xula.edu) and Megan N. Griff

Xavier University of Louisiana, New Orleans

Project Goals: The practical goal is to develop a new and powerful approach to quantitatively examine the dynamic behavior of bacterial electron transport systems at the microbe-mineral interface, an interaction that has heretofore been difficult to probe directly. We will accomplish this goal by refining and exploiting a novel integrating cavity absorption meter (ICAM) where the sample observation cell is also a spherical reflecting cavity. Light scattering losses due to sample turbidity will be eliminated or minimized by multiple reflections at the cavity wall that will prevent the scattered light from escaping the detector. The utility of this novel ICAM to study *in situ* bacterial electron transfer reactions will be demonstrated in real time equilibrium and kinetic measurements of electron transfer reactions in intact bacteria that respire by exchanging electrons with soluble and insoluble extracellular iron. The ability to conduct accurate real-time absorbance measurements in live organisms will permit any biological process that involves colored biomolecules to be studied in new ways.

This abstract introduces a new means to study respiratory electron transfer reactions *in situ* in intact bacteria under physiological solution conditions. The premise is that accurate UV-visible spectroscopy of electron transfer reactions among colored biomolecules can be conducted in highly turbid suspensions if the live bacteria are irradiated in an isotropic homogeneous field of incident measuring light. Under those conditions, the absorbed radiant power is independent of scattering effects. We conducted equilibrium and kinetic studies on the Fe(II)-dependent reduction and O_2-dependent oxidation of cytochromes in intact *Leptospirillum ferrooxidans* at pH 1.7. We used a novel integrating cavity absorption meter where the cuvette comprised a reflecting cavity completely filled with the absorbing suspension. *L. ferrooxidans* was selected because it is only known to respire on one substrate, reduced iron.

The aerobic iron respiratory chain of *Leptospirillum ferrooxidans* was dominated by the redox status of an abundant cellular cytochrome with an absorbance peak at 579 nanometers in the reduced state. Intracellular cytochrome 579 was reduced within the time that it took to mix a suspension of the bacteria with soluble ferrous iron at pH 1.7. Subsequent oxidation of the reduced cytochrome appeared to be the rate-limiting step in the overall aerobic respiratory process. Steady state turnover experiments were conducted where the concentration of ferrous iron was less than or equal to that of the oxygen concentration. The concentration of the reduced cytochrome 579 at any time point was directly proportional to the velocity of product ferric ion formation. Further, the integral of the area of the reduced cytochrome accumulated over time was also directly proportional to the total concentration of ferrous iron in each reaction mixture. These kinetic data obtained using whole cells were consistent with the hypothesis that reduced cytochrome 579 is an obligatory steady state intermediate in the iron respiratory chain of this bacterium.

The direct and accurate observation of absorbance changes *in situ* in intact organisms is a useful complement to traditional reductionist approaches and recent advances in proteomic and transcriptomic studies. The colored prosthetic groups of most electron transport proteins comprise intrinsic spectrophotometric probes whereby transient changes in the oxidation-reduction state of the proteins may be monitored with great sensitivity. There is no better means to establish physiological relevance in a metabolic function than to directly observe the function as it occurs in the intact bacterium. The movement of electrons through electron transfer complexes is central to energy production in all living cells. The ability to conduct direct spectrophotometric studies under noninvasive physiological conditions represents a new and powerful approach to examine the extents and rates of biological events *in situ* without disrupting the complexity of the live cellular environment. Studies such as these should increase our fundamental understanding of biological energy transduction.

This research was supported by the Office of Science (BER), U. S. Department of Energy.

150

DOE JGI Plant Genomics Program

Jeremy Schmutz[1,2]* (jschmutz@hudsonalpha.org), Jarrod Chapman,[2] David Goodstein,[2] Jane Grimwood,[1,2] Uffe Hellsten,[2] Jerry Jenkins,[1] Gerald A. Tuskan,[2,3] Jim Bristow,[2] Kerrie Barry,[2] Daniel S. Rokhsar,[2] and **Edward M. Rubin**[2]

[1]DOE JGI HudsonAlpha Genome Sequencing Center, Huntsville, Ala.; [2]DOE Joint Genome Institute, Walnut Creek, Calif.; and [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Project Goals: The goal of the DOE JGI Plant Genome Program is to shed light on the fundamental biology of photosynthesis and transduction of solar to chemical energy. Other areas of interest include characterizing:

1. **Ecosystems and the role of terrestrial plants and oceanic phytoplankton in carbon sequestration.**
2. **The role of plants in coping with toxic pollutants in soils by hyper-accumulation and detoxification.**
3. **Feedstocks for biofuels, e.g., biodiesel from soybean; cellulosic ethanol from perennial grasses and trees.**
4. **The ability to respond to environmental change (e.g., loss of diversity associated with changes in tempera-**

ture or moisture availability; nitrogen fixing nodules in legumes reduce fertilizer need).

5. **The generation of useful secondary metabolites for positive/negative pest control in natural ecosystems with attendant influence on global carbon cycle.**

The Plant Genome Program contributes to and accomplishes these goals through the following activities:

Sequence. Produce genome sequences of key plant (and algal) species to accelerate biofuel development and understand response to climate change.

Function. Develop datasets (and synthetic biology tools) to elucidate functional elements in plant genomes, with special focus on handful of "flagship" genomes.

Variation. Characterize natural genomic variation in plants (and their associated microbiomes), and relate to biofuel sustainability and adaptation to climate change.

Integration. Provide a centralized hub for the retrieval and deep integrated analysis of plant genome datasets.

JGI plant projects are initiated directly from three major sources: DOE facilities such as the BioEnergy Research Centers; the DOE scientific research communities via the JGI Community Sequencing Program; and JGI Plant Program Projects directed at the improvement of community resources for DOE plant science. These projects fall into broad scientific categories including: plant *de novo* genome sequencing and improvement, diversity and population analysis, transcript profiling with RNA-seq, and mapping and recombination analysis of populations.

151

Genome Improvement of DOE JGI Flagship Plant Genomes

David Sims[1]* (dsims@hudsonalpha.org), Jerry Jenkins,[1] Wellington Muchero,[2] Gerald A. Tuskan,[2,3] Jim Bristow,[3] Kerrie Barry,[3] Daniel S. Rokhsar,[3] Jeremy Schmutz,[1,3] Jane Grimwood,[1,3] and **Richard M. Myers**[1]

[1]DOE JGI HudsonAlpha Genome Sequencing Center, Huntsville, Ala.; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]DOE Joint Genome Institute, Walnut Creek, Calif.

Project Goals: As part of the DOE JGI Plant Program, we are working to experimentally and computationally improve the genomic sequences of key DOE plant species which we have designated the Plant Flagships. These genomic sequences and related data support mission aims in producing cellulosic feedstocks and understanding carbon cycling and carbon sequestration.

Through discussion with the JGI Plant Genome Advisory Committee and DOE, a set of JGI plant genomes that are the most important to DOE mission and plant science have been designated as JGI Plant Flagship Genomes. This selected set of critical species allows us to focus our computational and experimental efforts to move beyond sequence to function and to provide the most direct benefit to mission science. The current Plant Flagships are:

- Poplar—the DOE tree, the basis for cellulosic research at ORNL
- Sorghum—widely planted grass crop for biomass, cellulose, and sugar
- *Brachypodium*—small grass model organism
- *Chlamydomonas*—the most studied algal species, model algal organism
- Soybean—the source of biodiesel and the number two US economic crop
- Foxtail millet—a grass model, recently evolutionary diverged from switchgrass
- *Physcomitrella*—moss model organism, basic comparator for land plants

Although the flagship designation was initially intended for genomes that had already been draft sequenced, we also consider several additional genomes as proto-flagships because of their importance to DOE biofuel mission; initial sequencing efforts are on-going along with the creation of genomic resources:

- *Panicum virgatum* (switchgrass)—a candidate biofuel feedstock that grows on marginal soil and is being used by all of the BioEnergy centers as a model crop species
- Miscanthus—a perennial grass species that produces large amounts of cellulosic material with low agricultural inputs, one of the top feedstock candidates
- *Panicum hallii* (Hall's panicgrass)—a small, evolutionary nearby diploid relative of switchgrass that may serve as laboratory model organism for switchgrass research

For genome improvement, the objective is to completely (or as near as possible) resolve each base pair within a localized region and then using mapping resources, position and incorporate the sequence into chromosome scale pseudomolecule assemblies. In some cases, we have needed to construct the mapping resources ourselves to push forward a chromosome scale release. For base pair level improvement, which we carry out for the entire gene space of an organism or directed regions, we collect experimental data in the form of primer walks with a variety of chemistries and templates, transposon mediated sequencing and shatter libraries from clones (the latter two techniques being essentially subprojects of the subprojects). We have ongoing efforts to explore practical uses of next-generation sequencing technologies for improving the extremely large and polyploid plant genomes. We release new versions of the Plant Flagship genomes periodically as we complete significant iterative improvements. The sequence and subsequent annotation is made public through the phytozome web site at www.phytozome.net.

We have recently completed major updates to the sequences of *Chlamydomonas*, *Physcomitrella* and poplar, all of which are currently in the annotation phase. We are proceeding with a directed improvement project of sorghum gene space

and a complete improvement project for *Brachypodium*. With these improved sequences we seek to 1) provide near complete information for genomes of direct DOE mission importance, 2) support BRC and others efforts for cellulosic biofuel development and plant customization efforts, 3) foster communities to develop scientific research programs working with DOE organisms and attract new scientists to solve DOE problems by providing tool sets to work with these organisms, and 4) build a solid foundation for diversity, phenotype, resequencing, transcriptomic studies, and functional studies in these organisms.

152

Fidelity and Dynamics of DNA Methylation in Plants

Qin Yao[1]* (qyao@bnl.gov), Changjun Liu,[1] John Shanklin,[1] Chuan He,[2] and **John Dunn**[1]

[1]Department of Biology, Brookhaven National Laboratory, Upton, N.Y. and [2]Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Chicago, Ill.

Project Goals: The main goal of our program is to identify fundamental pathways and protein-DNA interactions maintaining inheritance of epigenetic states of plant DNA and chromatin during cell divisions. During semi-conservative DNA replication, hemi-methylated sites are generated where only cytosines in the parental strands remain modified. The correct methylation patterns in the daughter strands need to be reestablished to ensure proper epigenetic homeostasis, i.e. expression of essential genes and silencing of parasitic transposons. Recently we found that Arabidopsis genomic DNA contains 5-hydroxymethylcytosine (5hmC) as well as 5-methylcytosine (5mC)(1) which raises several questions: 1) where are these residues located; 2) do proteins that help maintain proper DNA methylation patterns in replicated DNA also maintain these patterns after replication across sites with 5hmC; and 3) do these two modifications play distinct epigenetic roles? To help address these questions, we are using a sensitive chemical labeling method to capture DNA fragments with 5hmC to discover its distribution pattern in the *Arabidopsis* genome and *in vitro* DNA binding studies to determine whether proteins needed for high-fidelity maintenance of cytosine methylation lose their intrinsic preference for hemi-methylated DNA when 5mC is replaced by 5hmC. Ultimately, an efficient and single-base resolution method for 5hmC identification is highly desirable. To this end we will integrate J binding protein, which has been shown to recognize single 5hmC residues(2) with next generation sequencing techniques to develop a new high-resolution 5hmC mapping tool which can be easily operated in a standard molecular lab. This project will also attempt to use genome-wide association analysis to enhance our understanding that any observed changes in DNA methylation status are non-random, gene-specific alterations that can result in gene activation or inactivation, especially for genes involved in lignocellulose biosynthesis and oil formation. All the newly generated epigenetic data and tools will be deposited into the BESC Knowledge Base (KBase).

In *Arabidopsis*, several different mechanisms are utilized to reestablish appropriate epigenetic methylation marks in the daughter strand cytosines. One of these mechanisms involves the five member VIM (Variant in Methylation) family of proteins which help reestablish daughter strand cytosine methylation at symmetric CG dinucleotides. Using recombinant proteins and modified double-stranded deoxyoligonucleotides, we observed that full-length VIM1 binds preferentially to hemimethylated DNA with a single modified 5mCG site; a result consistent with its known role in preserving DNA methylation in vivo following DNA replication. However, when 5hmC replaces one or both cytosine residues at a palindromic CpG site, VIM1 binds with approximately 15-fold lower affinity. VIM3, another member of the five member VIM family, has a single amino acid change, S317A, relative to VIM1, 2, 4 and 5 in the region known to be important for binding to methylated DNA. Changing this residue in VIM1 did not significantly affect its DNA-protein binding properties. These results suggest that 5hmC may contribute to VIM-mediated passive loss of cytosine methylation in vivo during *Arabidopsis* DNA replication. Work is now in progress to extend these observations to other proteins involved in maintaining epigenetic fidelity to determine their ability to discriminate between substrates with 5mC vs. 5hmC using purified recombinant proteins as well as protein binding microarrays (PBMs)(3).

Previous studies of DNA hydroxymethylation suggest hydroxylation of 5mC may promote transcriptional de-repression by dissociation of mC-binding proteins and/or recruitment of effector proteins(4). Our preliminary profiling of 5hmC residues in *Arabidopsis* reveals that 5hmC is enriched selectively in gene body regions rather than more distal regions and that there is a significantly different pattern around transcription start regions (TSS). This relative distribution is associated with gene expression levels. There is a highly negative correlation between enrichment in proximal regions to TSS's but no significant correlation between enrichment in the gene bodies and gene expression level. These results suggested that conversion of 5mC to 5hmC, possibly coupled to active or passive demethylation, might act as a switch to fine tune epigenetic homeostasis or perhaps have an active role in reprogramming the epigenome. Functional epigenomic studies will eventually address the mechanism from epigenotype to phenotype.

References

1. Yao, Q., Song, C.X., He, C., Dunn, J.J. Detection of 5-hydroxymethylcytosine in *Arabidopsis* and its potential impact on recognition of VIM1 (under review).
2. Robertson, A.B., Dahl, J.A., Vagbo, C.B., Tripathi, P., Krokan, H.E. and Klungland, A. (2011) A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res, 39, e55.

3. M.L. Bulyk, Analysis of sequence specificities of DNA-binding proteins with protein binding microarrays. (2006) Methods Enzymol 410, 279-299.
4. Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X. et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol, 29, 68-72.

153

PET Radiotracer Imaging in Plant Biology

Drew Weisenberger* (drew@jlab.org)

Thomas Jefferson National Accelerator Facility, Newport News, Virginia

The ability to detect the emissions of radioactive isotopes through radioactive decay (e.g. beta particles, x-rays and gamma-rays) has been used for over 80 years as a tracer method for studying natural phenomena. The Hungarian-born chemist George de Hevesy in 1913 while in Vienne published with Fritz Paneth the first account of the use a radioactive isotope as a tracer in a paper titled: "The Solubility of Lead Sulphide and Lead Chromate." In 1923 Hevesy published his studies on the transport of the radioisotopes lead-210 and lead-212 in living plants. In 1948 in a University of California Radiation Laboratory report Melvin Calvin and Andrew Benson describe research into photosynthesis of brown algae using the radioactive tracer carbon-14. More recently a positron emitting radioisotope of carbon: carbon-11 has been utilized as a $^{11}CO_2$ tracer for plant ecophysiology research. Because of its ease of incorporation into the plant via photosynthesis, the $^{11}CO_2$ radiotracer is a powerful tool for use in plant biology research. Positron emission tomography (PET) imaging has been used to study carbon transport in live plants using $^{11}CO_2$. Presently there are several groups developing and using new PET instrumentation for plant based studies. Instrumentation originally developed for small animal PET has been modified for imaging ^{11}C in plant research as currently undertaken at Institute Phytosphere, Forschungszentrum Jülich in Germany. A two-headed planar PET plant imaging system for ^{11}C has also been described and used in plant studies by a research group at the Japan Atomic Energy Agency. At Brookhaven National Laboratory researchers have used a clinical human PET scanner to track the distribution of ^{11}C labeled metabolites in plants in response to environmental changes. The group I head at Thomas Jefferson National Accelerator Facility (Jefferson Lab) in collaboration with the Duke University Phytotron is also involved in PET detector development for plant imaging. In this presentation I will provide a brief overview of radiotracer imaging in plants and review some of the latest developments of the use of $^{11}CO_2$ tracer in plant studies of others as well as the developments of the Jefferson Lab/Duke University effort.

154

Pan-omics Informatics and Advanced Analysis Pipelines

Gordon A. Anderson* (gordo@pnnl.gov), Ronald J. Moore, Joshua N. Adkins, Kenneth J. Auberry, Mikhail E. Belov, Kristin E. Burnum, Kevin L. Crowell, Stephen J. Callister, Therese R.W. Clauss, Kim K. Hixson, Gary R. Kiebel, Brian L. LaMarche, Mary S. Lipton, Da Meng, Thomas O. Metz, Matthew E. Monroe, Heather M. Brewer, Carrie D. Nicora, Angela D. Norbeck, Daniel J. Orton, Ljiljana Paša-Tolić, Samuel O. Purvine, John D. Sandoval, Anuj Shah, Yufeng Shen, Anil K. Shukla, Mudita Singhal, Gordon W. Slysz, Aleksey V. Tolmachev, Nikola Tolić, Karl Weitz, Aaron T. Wright, Rui Zhang, Rui Zhao, and **Richard D. Smith** (rds@pnnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

As part of our Proteomics and Pan-Omics project in support of GSP Systems Biology Research, we are developing an advanced informatics framework to enable application of pan-omics measurements that afford the comprehensive global molecular characterization necessary for understanding, modeling, and potentially manipulating biological systems. Pan-omics measurement capabilities are based on essentially identical separations-MS measurement platforms and similar data processing/informatics pipelines: metabolomics, lipidomics, and glycomics measurements in addition to expanded proteomics measurements. A key element of the pan-omics informatics framework are tools and methodologies for integrating data from various omics measurements and determining data quality and ambiguities (e.g., the confidence in peptide and protein identifications, modification sites, and abundance levels), as well as approaches for managing and communicating data.

We are leveraging the extensive experience and capabilities developed to date at PNNL to extend high throughput proteomics measurements to multiple omics measurements and provide a framework for evaluating and controlling data quality; processing and integrating data from the various analysis streams; and disseminating data and information to collaborators, users, and the broader scientific community. The Pan-omics Research Informatics Storage and Management (PRISM) system has been enhanced to support advanced pan-omics analysis pipelines that allow automated high level data analysis and integration of hundreds to thousands of individual datasets into one comprehensive research result report. The framework further supports integration of genomics data from public repositories and aims to provide the needed infrastructure to interoperate with the GTL Knowledgebase.

Current efforts are focused on developing a suite of data analysis tools, data consolidation applications, and statistical packages, as well as visualization software for data inter-

pretation and integration. New and advanced tools support metabolomics, glycomics, and top-down proteomics pipelines, as well as pan-omics data integration. Each pipeline produces a self-documented, tested, and repeatable analysis, complete with visual inspection plots produced for each step in the pipeline. The advanced automation and critical infrastructure elements demonstrate high throughput and automated file processing, as well as data filtering and result aggregation.

Acknowledgements: This research is supported by the Office of Biological and Environmental Research (OBER) of the U.S. Department of Energy (DOE). Portions of this research were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE's OBER. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

155

Proteomics and Pan-omics Measurements for Comprehensive Systems Characterization of Biological Systems

Mary S. Lipton* (mary.lipton@pnnl.gov), **Thomas O. Metz*** (thomas.metz@pnl.gov), Kristin E. Burnum, Carrie D. Nicora, Qibin Zhang, Stephen J. Callister, Scott R. Kronewitter, Aaron T. Wright, Ronald J. Moore, Zeping Hu, Xiaoli Gao, Weijun Qian, Erin S. Baker, Therese R.W. Clauss, Karl K. Weitz, Daniel J. Orton, Matthew E. Monroe, Anil K. Shukla, Rui Zhao, Roslyn Brown, Joshua N. Adkins, **Gordon A. Anderson**, and **Richard D. Smith** (rds@pnnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: This project is implementing and applying advanced capabilities for comprehensive molecular characterization of biological systems, including the extension of proteomics to cover post-translational protein modifications and the implementation of broad metabolomics, lipidomic and glycomic measurements, together with more widely available genomics and transcriptomics capabilities. Applications to microbes, plants and microbial communities enhance scientific understanding by elucidating phenotypic relationships between environmentally important microorganisms and higher organisms, as well as metabolic activities within microbial communities.

The goal of BER's Genome Science Program (GSP) is to achieve a predictive systems level understanding of plants, microbes and biological communities via the integration of fundamental science and technology developments to enable biological solutions to challenges in energy, environment and climate. Achieving this goal requires comprehensive proteomics, metabolomics, lipidomics, and glycomics, i.e. pan-omics measurement capabilities, and the integration of data generated by these approaches. We are facilitating understandings of biological systems by applying pan-omics molecular measurement capabilities in biology-driven collaborative projects led by investigators actively engaged in developing systems biology approaches in support of BER's research agenda. Our strategy benefits from advances in high resolution nano-liquid chromatography (LC) separations combined with high mass accuracy mass spectrometry (MS) measurements and other developments that afford large gains in measurement quality and throughput. These efforts also include the automation of key steps in proteomics sample processing; fractionation of protein samples based on surface membrane protein enrichment and subcellular fractionation methods using differential gradient centrifugation; and implementation of novel methods for protein extraction from environmental (e.g., soil) samples. Additional advancements involve the implementation of targeted proteomics methods (e.g., activity-based protein profiling and multiple reaction monitoring) and approaches for elucidating protein isoforms (e.g., integrated top-down and bottom-up proteomics) and post-translational modifications (e.g., phosphoproteomics and characterization of protein glycosites).

The application to studies of fungus-growing ant–microbe symbiosis as a paradigmatic example of organic complexity generated through symbiotic association illustrates integration of these capabilities. Over the last decade, it has become a model system for studying symbiosis. We have demonstrated in-depth profiling of the fungal garden complete with bacteria (fungus alone, isolated bacteria, and fungal garden intact) to understand the relationship between the fungus and the bacterial protectors. Proteomics and metabolomic studies of the secreted proteins from the bacteria have been characterized in an effort to understand the relationship between the ants and the fungus. These studies demonstrate the ability to use pan-omics measurements on an ecosystem level, spanning bacteria to multi-cellular organisms.

Acknowledgements: This research is supported by the Office of Biological and Environmental Research (OBER) of the U.S. Department of Energy (DOE). Portions of this research were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE's OBER. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

156

Initial Application of a New LC-IMS-TOF MS Platform to Complex Environmental Samples

Erin Shammel Baker* (erin.baker@pnnl.gov), Yehia M. Ibrahim, William F. Danielson III, Kristin E. Burnum, Matthew E. Monroe, Kevin L. Crowell, Daniel J. Orton, Gordon W. Slysz, Brian J. Anderson, Carrie D. Nicora,

Mary S. Lipton, Gordon A. Anderson, and **Richard D. Smith** (rds@pnnl.gov)

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.

As part of our Proteomics and Pan-Omics project in support of GSP Systems Biology Research, we are developing advanced pan-omics measurements capabilities for comprehensive global biomolecular characterization to enable the modeling, understanding, and potential manipulation of complex biological systems. Development and application of advanced separations-MS measurement platforms greatly increase measurement quality and throughput. The new platform combines fast, multidimensional separations (i.e., fast LC in conjunction with millisecond-scale ion mobility separations) with ultra-fast and accurate mass measurement time-of-flight MS to greatly expand biomolecule coverage and sensitivity, and the addition of the third dimension ion mobility separation allows for differentiation between molecule types. Pan-omics measurement capabilities are based on essentially identical separations-MS measurement platforms and similar data processing/informatics pipelines for metabolomics, lipidomics, and glycomics measurements, as well as for expanded proteomics measurements that cover a range of important modifications.

The extreme complexity of environmental samples presents many challenges for currently available MS-based analytical platforms. High concentrations of organic material (such as humic acid in soils and polyphenols in leaves) and sources of contamination from natural environments (such as high salts or other minerals) make environmental samples extremely difficult to analyze. Our new platform that utilizes advanced separations in conjunction with MS greatly increases measurement quality and throughput, even for "dirty" environmental samples. The new platform utilizes fast LC and an ion mobility separation (IMS) coupled with an ultra-fast and accurate mass measurement time-of-flight (TOF) MS to provide expanded proteome coverage and greater sensitivity when high concentration species are present in complex samples.

Previous studies of environmental ocean water, soil and leaf samples that have employed more conventional trapping MS instruments (e.g., LTQ Orbitrap MS) have presented a number of analytical challenges. For example, in peptide-centric analyses of complex samples, large collections of ionized species regardless of the type of molecule are essentially analyzed simultaneously in the ion trap that has a finite capacity. High concentrations of non-bio-molecular species eluted for extended periods of time rapidly fill the trap, making for very short accumulation times, which results in the inability to detect many lower concentration species that co-elute with the high concentration species. TOF mass spectrometers offer a promising solution to help with environmental samples in that they do not have an ion trap, so all ions are sampled equally at the detector instead of being limited by co-eluting high concentration species. These platforms also allow effective utilization of advanced, highly efficient ion sources and interface designs (e.g., using ion funnels) that increase dynamic range and detection of lower abundances species. Modern TOF mass spectrometers are also capable of both very high resolution and excellent (e.g., low ppm) mass measurement accuracies. The greatest advantage afforded by the new platform is that IMS is able to separate different classes of molecules onto specific 'trend lines' (e.g., peptides, lipids, oligonucleotides, etc.) because of distinctive backbones that make the molecules drift differently through the buffer gas and vastly different from small single ions or larger branched organic molecules. As a result, it is possible to separate and distinguish environmental contaminants from peptides or other biomolecule components in the same analysis.

Samples whose analyses have been known to be problematic with conventional MS-based platforms, such as ocean water isolates, soil extracts with humic acids, and plant extracts with poly phenols have recently been analyzed using the LC-IMS-TOF MS platform that provided coverage and measurement dynamic range much greater than with trapping-based MS platforms. Additionally, peptides and other biomolecules that have altered structures and different drift patterns can be distinguished, e.g., peptides with different post-translational modifications, adducts, or those that have been cross-linked. Modifications and adducts occurring in the same type of molecules also can be readily distinguished with the LC-IMS-TOF MS platform. For example, we recently found that mercury-modified peptides travel through the IMS buffer gas faster than peptides of the same mass-to-charge ratio, while phosphor- and heme-containing peptides travel slower because of the conformational changes induced by the modifications and adducts. By knowing these trends, specific modifications can be observed from complex mixtures.

This poster will highlight several complex environmental samples examined with commercially available instruments and with the IMS-TOF MS platform to illustrate the improvements in sensitivity and coverage observed with the new platform.

Acknowledgements: This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DEAC05-76RLO 1830.

157

Integrating Transcriptomics and Proteomics Data for Identifying Mechanisms Regulating the Response of Microbes to Environmental Change

H. Steven Wiley[1]* (steven.wiley@pnnl.gov), Ronald C. Taylor,[2] Meng L. Markillie,[3] Mary S. Lipton,[3] Gretta Serres,[4] and Margie Romine[3]

[1]Environmental Molecular Sciences Laboratory, [2]Computational Biosciences, [3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and [4]Marine Biological Laboratory, Woods Hole, Mass.

Project Goals: The general goal was to identify global regulatory strategies used by microbes when responding to environmental changes. By integrating simultaneous proteomics and transcriptomics measurements, we sought to determine which cellular processes were regulated at the transcriptional versus post-transcriptional level. Using *Shewanella* as an initial test case, we developed approaches for accurately quantifying the level of protein and mRNA expression, integrating these data sets and visualizing their relationships on a global level. This approach was then evaluated for its ability to identify novel regulatory mechanisms and to obtain insights into regulatory strategies used by microbes in responding to environmental changes.

There are numerous mechanisms that regulate the response of microbial communities to changes in their environment. These include changes in the community composition as well as alterations in the complement of proteins that dictate the behavior of individual cells. Most investigations on how individual cells are regulated have focused on transcriptional control mechanisms, but there are other levels that could be important, such as translational control and protein turnover. The lack of suitable technologies to screen for alternate regulatory mechanisms on a global level has made it difficult to determine how common alternate mechanisms are and the role they might play in regulating microbial responses. To address this need, we developed a general and global approach for discriminating transcriptional from post-transcriptional regulatory mechanisms. The basic approach is to simultaneously quantify and compare the levels of proteins and mRNA transcripts on a global level. Proteins whose levels change in concordance with changes in their mRNA transcripts are considered transcriptionally regulated. Those that show discordance between changes in mRNA and protein levels are potentially regulated at a post-transcriptional level.

As a test case, we used *Shewanella* grown at steady state in chemostats under either high or low oxygen conditions. Samples were collected on three different days and subjected to quantitative proteomics and transcriptomics analysis. For proteomics, quadruplicate technical replicates were run for all three sets of biological replicates and peptides identified using the AMT approach (accurate mass and time). Transcriptional profiling was performed using SOLiD sequencing with triplicate technical replicates. The proteomics data were normalized and converted to relative protein levels by combining average ion intensity measurements and peptide counts followed by corrections for protein length. The mRNA levels of each measured protein was normalized and corrected for gene length. The average protein:mRNA ratio for each gene was then calculated for each condition. To visualize changes in the protein:mRNA ratios for these extremely large data sets, we created a "double ratio" plot where the mRNA ratio for each gene under the two treatment conditions was plotted against the ratios of their protein:mRNA ratios. These data were plotted together with information on peptide counts and ion intensity measurements using the exploratory data analysis program Aabel. This allowed us to readily identify proteins that displayed evidence of posttranscriptional regulation.

The absolute amounts of protein and mRNA expressed in *Shewanella* showed a general correlation at the level of individual genes (C.C. between 0.35-0.37). Binning the data into 25 equal groups ranked by absolute levels of gene expression showed much stronger correlation (C.C. ~0.98), suggesting that in general, the most abundant transcripts give rise to the most abundant proteins. Comparing changes of mRNA and protein expression in response to high and low oxygen conditions yielded similar results (C.C. ~0.43), thus supporting the hypothesis that the dominant mode of regulating protein levels is through transcriptional control. Nevertheless, there were several hundred proteins that showed evidence of posttranscriptional regulation. For example, the relative expression level of ribosomal proteins was essentially invariant despite changes in transcript abundance for several of them. Other proteins involved in regulating mRNA translation displayed a similar pattern. A clear case of post-transcriptional regulation was observed in the operon (*hypAEDCB*) which encodes the NiFe hydrogenase maturation complex. During high oxygen conditions, the mRNA of this operon goes down by about 2-3 fold. However, the corresponding protein levels go down 15-40 fold. Proteins from flanking operons or those that encode the NiFe hydrogenase showed no disparity between transcript and protein levels, suggesting that translation of the polycistronic message encoding HypAEDCB is specifically regulated. Preliminary pathway enrichment analysis of proteins displaying evidence for posttranscriptional regulation shows enrichment of leucine biosynthesis (p-value=1.8e-4), formate to TMAO electron transfer/anaerobic respiration (p-value=0.0023/0.013) and nitrogen metabolism (p-value=0.0085).

Our studies suggest that although most microbial proteins are regulated at the level of transcription, a significant fraction is regulated at the posttranscriptional level, particularly those involved in translation itself. This highlights the needs to include proteomics measurements in the analysis of microbial systems. The integrated approach we have developed for identifying the level at which proteins are regulated should also prove useful in the analysis of more complex microbial systems.

* Presenting author

This research was performed in the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory (PNNL) and is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the PNNL Foundational Scientific Focus Area.

Early Career Award

158

Quantitative Proteomic Profiling of Site-Specific Redox Modifications, Proteolytic Processing, and N-glycosylation

Dian Su,[1] Jong-Seo Kim,[1] Lu Wang,[1] Uma K. Aryal,[1] Yi Qu,[1] Aaron T. Wright,[1] T.C. Squier,[1] Mary S. Lipton,[1] Richard D. Smith,[1] Ziyu Dai,[2] Scott E. Baker,[2] and **Wei-Jun Qian**[1]* (weijun.qian@pnnl.gov)

[1]Biological Sciences Division and [2]Energy and Efficiency Division, Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: One of the primary objectives of this early career research project is to develop novel proteomic approaches to enable quantitative measurements of site-specific regulatory protein posttranslational modifications (PTMs) in subcellular compartments. The ability to effectively and quantitatively characterize site-specific PTMs is essential for understanding the regulation of cellular signaling and protein functions at the post-translational level, and for enabling a systems biology approach to study organisms important for bioenergy or environmental applications. Our efforts have been primarily focused on three important classes of PTMs: (1) reversible redox modifications on cysteinyl thiols, (2) proteolytic processing and protein N-terminal modification, and (3) glycosylation. All three classes of modifications are ubiquitous in both prokaryotic and eukaryotic cells and their important roles in cellular regulation and signaling have been increasingly recognized.

Reversible redox modifications of protein thiols: Functional cysteinyl residues in proteins serve as "redox switches" through reversible oxidation, which is recognized as a fundamental mechanism of redox regulation in almost all organisms. We are developing approaches to enable parallel identification and quantification of several types of reversible thiol modifications such as S-nitrosylation (SNO) in response to different cellular reactive oxygen species. The strategy for achieving site-specific proteomic identification is illustrated in **Fig. 1A.** Briefly, thiol specific modifications can be reduced by specific reagents and the converted free thiols are subsequently captured and enriched by a thiol-specific resin and their dynamics can be quantified by isobaric labeling and LC-MS/MS. We have applied this approach to profile SNO in *Synechococcus* sp. PCC 7002. We identified 226 cysteine residues sensitive to SNO modification (**Fig. 1B**) and many of the sites were identified in proteins localized to the thylakoid, indicative of the role of redox regulation in photosynthesis. Similarly, the application of this approach to a fungal organism, *Aspergillus niger*, also resulted in the identification of ~700 SNO-modified sites with many of the proteins being metabolic enzymes.

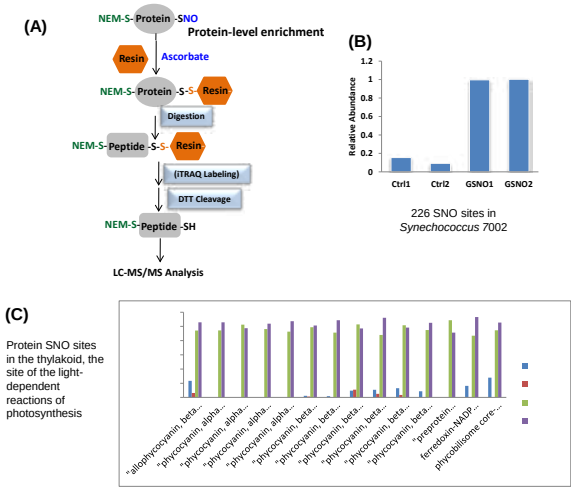


Figure 1. (A) The strategy for identification and quantification of S-nitrosylation (SNO); (B) SNO sites identified in 7002; (C) SNO sites identified in the thylakoid of 7002.

Proteolytic processing and N-terminal modifications: Proteolytic processing and *in vivo* N-terminal modifications are another common mechanism of regulation of protein functions. We have developed two complementary approaches to target specific protein N-termini and enrich N-terminal peptides for proteomic identifications. Applying these approaches to a whole cell lysate of *Aspergillus niger*, we were able to identity more than 2000 N-terminal sites with 228 *in vivo* modifications from 858 genes (**Fig. 2**). The number of N-terminal sites identified per gene reveals the complexity of the proteome due to proteolytic processing, which may reveal many novel functional proteins. Quantitative profiling will be performed to identify functional products.

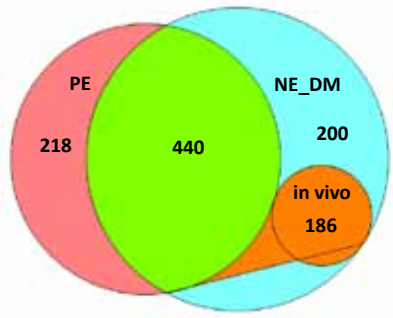


Figure 2. Number of genes identified with N-terminal sites. PE, positive enrichment; NE, Negative enrichment; *in vivo,* proteins with N-terminal sites modified *in vivo.*

Protein glycosylation: N- and O-linked glycosylation are known to play an essential role in cellular functions and secretory pathways. We have adapted and optimized of a hydrazide chemistry based enrichment approach for profiling N-glycosites. We have performed an extensive mapping

of N-glycosylated sites in *A. niger* by applying this approach using hydrazide-modified magnetic beads. The optimized protocol was applied to profile N-glycosylated sites from both the secretome and whole cell lysates of *A. niger*. A total of 847 N-glycosylated sites from 330 N-glycoproteins (156 proteins from the secretome and 279 proteins from whole cells)[1] were confidently identified by LC-MS/MS. The identified N-glycoproteins in the whole cell lysate were primarily localized in the plasma membrane, endoplasmic reticulum, Golgi apparatus, lysosome, and storage vacuoles (**Fig. 3**), supporting the important role of N-glycosylation in the secretory pathways. The extensive coverage of N-glycosylated sites and the observation of partial glycan occupancy on specific sites in a number of enzymes provide important initial information for functional studies of N-linked glycosylation and as a prelude to their biotechnological applications in *A. niger*.

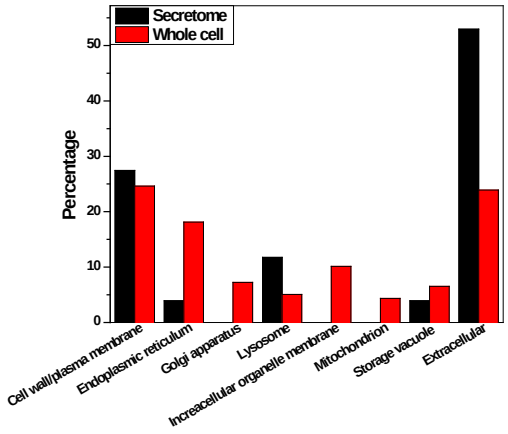


Figure 3. N-glycoproteins identified in cellular compartments based on gene ontology information.

Reference

1. Wang, L., Aryal, U. K., Dai, Z., Mason, A. C., *et al.*, Mapping N-Linked Glycosylation Sites in the Secretome and Whole Cells of Aspergillus niger Using Hydrazide Chemistry and Mass Spectrometry. *J. Proteome Res.* 2011.

This work is supported by a DOE Early Career Research Award under the Office of Biological and Environmental Research in the DOE Office of Science. Mass spectrometry-based proteomics measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

159

Optimization of Microbial Community Proteomics

W.R. Cannon[1]* (william.cannon@pnl.gov), A. Hugo,[1] D.J. Baxter,[2] A. Kalyanaraman,[3] and B.-J. Webb-Robertson[1]

[1]Computational Biology and Bioinformatics Group, [2]Molecular Sciences Computing Facility, Pacific Northwest National Laboratory, Richland, Wash.; and [3]School of Electrical Engineering and Computer Science, Washington State University, Pullman

Project Goals: Analyzing proteomics data from environmental samples is extremely challenging. Generally, the analyses fall into one of two categories, either the complete lack of metagenome sequence information to match proteome data against, or the incomplete nature of metagenome sequence information. The goals of this 3-year SciDAC/GSP project were to alleviate this situation by leveraging high performance computing, advanced modeling and algorithms to develop more accurate and more complete analyses. Achievements 1-3 below address the accuracy issue, and achievements 4-6 address issues regarding complete analysis of metaproteome/metagenome data.

1. **Developed advanced models based on the mathematical overlap between statistical data analysis and statistical thermodynamics [1].**
2. **Implemented the advanced models to bear on problems relevant to DOE's mission in Bioenergy [2].**
3. **Released code for these tools that can be used on multiple computing platforms including workstations, HPC clusters and cloud resources [2-4].**
4. **Developed HPC solutions to the challenge of detecting protein sequence homology for expanding genome and metagenome datasets [5, 6].**
5. **Developed methods to analyze samples when no reference genomes or metagenomes exist [7].**
6. **Developed methods to maximally use metagenome sequences when uncertainty about the gene calls is not small [8].**

These last four achievements are discussed briefly below.

Proteotyping Environmental Samples without a Metagenome. When a metagenome sequence is not available, proteomic analysis is often not possible. This year we reported the development of a novel high performance computing method for proteotyping environmental samples. The method uses computational optimization to provide an effective way to control the false discovery rate. The method provides phylum/species information based on the expressed proteins in a microbial community, and thus complements DNA-based methods. Testing on blind samples demonstrates that the method provides 79-95% overlap with analogous results from searches involving only the correct genomes. Scaling and performance evaluations for the software demonstrates the ability to carry out large-scale

optimizations on 1258 genomes containing 4.2M proteins [7].

Proteome Matching against a Metagenome: Bayesian Integration of Evidence. When a metagenome sequence is available, specific identifications of proteins are possible. However, due to the imperfect nature of metagenome sequence information, the protein identification rate is usually quite a bit lower than what is typical for single organism laboratory studies. This need not be the case if multiple levels of evidence are assessed when identifying proteins from an environmental sample. We report the development of a Bayesian framework that incorporates evidence based on peptide detectibilities, significance of MSMS matches, and prior probabilities of protein occurrence. The method more than doubles the number of spectra that are identified with proteins, and increases the number of proteins that are identified by 40-50%.

Proteomics Analysis Code: MSPolygraph. The analyses above were carried out using high performance computing. MapReduce, MPI and serial versions of *MSPolygraph* for peptide identification from mass spectrometry data have been developed. The MPI version runs on any MPI-capable cluster [2]. The MapReduce implementation can run on any Hadoop cluster environment [3]. Availability: The source code along with user documentation is available at omics.pnl.gov/software/MSPolygraph.php.

Remote Homology Detection. Large-scale projects generate millions of *new* sequences that need to be matched against themselves and against already available sequences. For example, the ocean microbiota survey project in 2007 analyzed a total of 28.6M sequences. The most time consuming step during analysis was homology detection, which accounted for 10^6 CPU hours despite the use of fast approximation heuristics such as BLAST.

We developed a novel parallel algorithm, *pGraph,* to efficiently parallelize the construction of sequence homology graphs from large-scale protein sequence data sets based on dynamic programming alignment computation. The parallel design is a hybrid of multiple-master/worker and producer-consumer models, which effectively addresses the unique set of irregular computation issues and input data availability issues. The implementation scales linearly up to 2,048 processors using up to 2.56×10^6 metagenomic protein sequences [6].

Computations were performed in the National Energy Research Scientific Computing Center (NERSC) in Berkeley, CA, and in the Molecular Sciences Computing Facility at the Environmental Molecular Sciences Laboratory (EMSL).

References

1. W. R. Cannon and M. M. Rawlins, "Physicochemical/Thermodynamic Framework for the Interpretation of Peptide Tandem Mass Spectra," *Journal of Physical Chemistry C,* vol. 114, pp. 5360-5366, Apr 1 2010.
2. W. R. Cannon, M. M. Rawlins, D. J. Baxter, S. J. Callister, M. S. Lipton, and D. A. Bryant, "Large Improvements in MS/MS-Based Peptide Identification Rates using a Hybrid Analysis," *J Proteome Res,* Mar 30 2011.
3. A. Kalyanaraman, W. R. Cannon, B. Latt, and D. J. Baxter, "MapReduce implementation of a hybrid spectral library-database search method for large-scale peptide identification," *Bioinformatics,* vol. 27, pp. 3072-3, Nov 1 2011.
4. G. Kulkarni, A. Kalyanaraman, W. R. Cannon, and D. Baxter, "A Scalable Parallel Approach for Peptide Identification from Large-Scale Mass Spectrometry Data," presented at the International Conference on Parallel Processing Workshops (ICPPW 09), 2009.
5. C. Wu, A. Kalyanaraman, and W. R. Cannon, "A scalable parallel algorithm for large-scale protein sequence homology detection," in *39th International Conference on Parallel Processing (ICPP)*, 2010, pp. 333-342.
6. C. Wu, A. Kalyanaraman, and W. R. Cannon, "pGraph: Efficient Parallel Construction of Large-Scale Protein Sequence Homology Graphs," *Revision submission- IEEE Transactions on Parallel and Distributed Systems,* 2012.
7. A. Hugo, D. J. Baxter, W. R. Cannon, A. Kalyanaraman, G. Kulkarni, and S. J. Callister, "Proteotyping of Microbial Communities using High Performance Optimization of Proteome-Spectra Matches," in *Pacific Symposium BioComputing*, In Press, 2012.
8. A. Hugo, B. J. Webb-Robertson, and W. R. Cannon, "Increased Proteome Coverage from Environmental Samples using Bayesian Analysis of Evidence," in *In Preparation*, 2012.

This research was supported by the Genomic Science Program, Office of Biological and Environmental Research, and the Scientific Discovery through Advanced Computing program, Office of Advanced Scientific Computing Research, U.S. Department of Energy.

160

PUNCS: Towards Predictive Understanding of Nitrogen Cycling in Soils

Robert Sanford[1]* (rsanford@illinois.edu), Allana Welsh,[1] Kostas Konstantinidis,[2] Joanne Chee-Sanford,[3] Kirsti Ritalahti,[4] Sukhwan Yoon,[4] and **Frank Löffler**[4,5]

[1]University of Illinois, Urbana-Champaign; [2]Georgia Institute of Technology, Atlanta; [3]USDA-ARS, Urbana, Ill.; [4]University of Tennessee, Knoxville; and [5]Oak Ridge National Laboratory, Oak Ridge

Project Goals: In anoxic environments, nitrate is largely consumed by dissimilatory nitrate reduction to ammonium (DNRA) and nitrate reduction to N_2O and N_2 (denitrification). DNRA oxidizes more C per mole of nitrate than denitrification and generates a cation (NH_4^+), which is retained in most soils, indicating that these processes have profoundly different impacts on N retention and greenhouse gas (CO_2, N_2O) emissions. Microbes capable of DNRA or denitrification coexist but the regulatory controls for these competing processes are unknown, and their relative contributions cannot be predicted faithfully with current models. This paucity of information limits the development of more accurate, predictive models of N-flux including the effects of N-retention on plant

growth and biomass yield for bioenergy production, as well as greenhouse gas (e.g., N_2O) emissions.

To elucidate the environmental factors controlling N-turnover in soils, the project team will conduct physiological studies, stable isotope probing (SIP) experiments, (meta)genomics and (meta)transcriptomics analyses, along with high resolution imaging (NanoSIMS) on systems of increasing complexity (i.e., pure cultures to mesocosms established with distinct soil types). We will quantitatively monitor genes and transcripts related to N-turnover in response to changes in pH, temperature, soil moisture, C- and N-content and assess the emitted gases using a mass balance approach based on (geo)chemical and stable isotope measurements. By comparing the gene-centric and genomic (*who is there?*) to the transcriptomic, SIP, and NanoSIMS (*who is how active?*) datasets gathered from different treatments, ***system-level insights into the pathway controls and the functional redundancy within microbial communities controlling N-flux in soils will be obtained***.

One project task determines the pathway controls for nitrate depletion in anoxic soil environments; denitrification versus DNRA. Previous observations suggested that the carbon to nitrogen (C:N) ratio controls the fate of nitrate to gaseous products (N_2O, N_2) or NH_4^+. *Shewanella loihica*, a unique bacterium with both complete denitrification and DNRA pathways, is used to explore the environmental factors (e.g., C:N ratio, type of C available, pH, temperature) that determine which N-oxide reduction pathway predominates. The results obtained to date suggest that the C:N ratio is one of several factors that influence the fate of nitrate. DNRA appears to be favored over denitrification under conditions of low nitrate:nitrite ratios and high C:N ratios. Further, nitrate reduction to ammonium predominated in mineral medium augmented with amino acids. Gene expression studies are being used to monitor the expression of DNRA and denitrification genes under these different growth conditions.

Another project goal is to better understand the diversity of *nosZ* genes involved in N_2O reduction to N_2. Genome analysis of the nitrite-to-NH_4^+-reducing, non-denitrifying bacterium *Anaeromyxobacter dehalogenans* strain 2CP-C revealed the presence of a complete *nosZ* gene cluster. Subsequent physiological studies corroborated that this organism uses N_2O as a growth-supporting electron acceptor. The denitrifier- and *Anaeromyxobacter*-types of *nosZ* genes share sequence similarity; however, the primers used for environmental surveys of denitrifier *nosZ* have failed to detect the *Anaeromyxobacter*-type *nosZ*. PCR primers specifically targeting the *Anaeromyxobacter*-type *nosZ* determined its distribution in different soil ecosystems, suggesting that an important, yet unrecognized N_2O sink exists. Recent isolation and sequencing efforts revealed an unexpected diversity of non-denitrifying N_2O reducers and *nosZ* genes in soil ecosystems suggesting that current N_2O emission models are missing a possible significant N_2O sink.

To elucidate the relative abundance of genes implicated in nitrate, nitrite and N_2O transformation in soils, we have embarked on surveys of existing metagenome datasets. The datasets incorporate millions of short-read sequences (e.g., shorter than 400 bp), obtained using either the Illumina or the Roche 454 sequencers. Therefore, the first objective of our work is to develop the bioinformatics pipeline that will allow us to reliably identify and align short fragments of the target genes recovered in the available metagenomes. Subsequently, we will evaluate in-silico the specificity of available primers for chosen target genes and design new primers for microbial groups that are not encompassed by the primers currently available. These approaches will enable the identification and enumeration of keystone microbial groups that respond to different incubation conditions and perturbations in mesocosm systems. We have also validated and applied Illumina Hi-Seq 2000 sequencing on soil samples, and have established the bioinformatics approaches to assemble such high-volume data (>50Gb per sample) and allow comparisons of different communities based on metagenomic datasets. We will report on recent developments including preliminary results from metagenome sequencing efforts.

This research is supported by DOE Genomic Science Grant DE-FG02-11ER65267.

161

Spatioelectrochemistry: The Molecular Basis for Electron Flow Within Metal-Reducing Biofilms

Daniel R. Bond[1]* (dbond@umn.edu), Ying Liu,[1] **Caleb Levar**,[1] and Janet Rollefson[2]

[1]Biotechnology Institute and Department of Microbiology and [2]Department of Biochemistry, Biophysics, and Molecular Biology, University of Minnesota, Twin Cities

Project Goals: Electrochemical, spectral, genetic, and biochemical techniques have provided evidence that multiple redox proteins and structural macromolecules outside the cell work together to move electrons long distances between *Geobacter* cells and to metals. This extracellular matrix contains many stable proteins not easily solubilized by standard methods, in addition to complex polysaccharides. Thus, much of these components were likely lost or ignored in previous proteomic and biochemical surveys. We aim to define this matrix and develop new tools to visualize it in action. The goals of this project are to 1) identify protein and polysaccharide elements crucial to both the assembly and function of the extracellular conductive matrix, 2) expand spectroelectrochemistry techniques to define the mechanism and route of electron transfer through the matrix, and 3) we will combine this knowledge of matrix proteins and their role in multicellular electron transfer to visualize redox and gene expression gradients in space over time.

* Presenting author

When bacteria change the state of metals in the environment, they transport electrons unprecedented distances from intracellular metabolic reactions to distant mineral surfaces. This electron movement drives subsurface bioremediation, controls aquifer chemistry, and powers new microbial energy generation applications. Yet a molecular understanding of how this electron transfer is accomplished by *Geobacteraceae,* who are among of the most predominant bacteria in such systems, remains one of the grand challenges in microbial environmental processes.

Direct measurements of living biofilms using electrochemistry has revealed electron hopping between *Geobacter* redox proteins to be a rate-controlling step at all stages of growth. Direct spectral analysis of living biofilms has confirmed that *c*-type cytochromes are a major reservoir of charge in these films, and that these cytochromes experience a bottleneck to oxidation when electrons must be transferred longer distances. Fine-scale immunogold labeling has discovered gradients in cytochrome abundance throughout these films, further suggesting the presence of gradients within this biofilms. Genetics has discovered new polysaccharide biosynthesis operons, and secretion systems essential for the attachment of *Geobacter* to metals and other cells in the biofilm. Biochemistry has shown that the extracellular space acts anchors a diverse assemblage of essential *c*-type cytochromes and adhesion proteins outside the cell. We hypothesize that this data converges on a model of electron transfer mediated by multiple cooperating proteins, ultimately limited by hopping between a highly adaptable network of cytochromes, which are attached to polysaccharides and pili.

Our project hypothesizes that it is the components of this matrix, and its high adaptability, that explains why *Geobacter* can so easily interface with a wide range of mineral surfaces, grow as multicellular networks, and quickly adapt to disruptions in single cytochromes.

To accomplish these goals, we are screening mutant libraries for strains defective in different developmental stages of this matrix using a combination of traditional and high-throughput approaches, and developing biochemical extraction methods for separating and analyzing this matrix. Key to this phase is to identify proteins required for specific developmental stages such as surface recognition, vs. self-recognition, or cell-surface electron transfer vs. cell-cell electron exchange. In parallel, we are inventing spectroelectrochemistry cells able to monitor cytochrome redox states in living biofilms, and engineering proteins for spatiotemporal localization of activity. This multidisciplinary approach aims to link quantitative data for specific reactions that occur at the biotic-abiotic interface with genes and expression patterns that can be used in predictive modeling, environmental monitoring, and design of bacteria with altered conductive properties.

162

Metabolic Dynamics of Starvation Induced Dormancy in *Saccharomyces cerevisiae*

Niels Klitgord[1]* (njklitgord@lbl.gov), Fredrik Alsin,[1] Theodore Peters,[2] Matt Youngblood,[1] Robert E. Hughes,[2] and **Trent Northen**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]Buck Institute, Novato, Calif.

Project Goals: To understand the metabolic mechanism involved in the initiation and departure of starvation induced dormancy in yeast.

Many organisms have adapted evolutionary to dealing with various stresses by entering a state of dormancy. Organisms that enter this dormancy typically make physiological changes that allows them to shut down their metabolism, and become resistant to various environmental perturbations including dessication, pH, and heat stresses. However the specific environmental, metabolic and physiological changes an organism undergoes as it enters and leaves dormancy is not well understood.

Here we use the yeast *Saccharomyces cerevisiae* to address the question of what metabolic changes occur as they enter a starvation induced dormancy. We attempt to answer this question by using gas-chromatography mass-spectrometry (GC-MS) to profile the intracellular and extracellular metabolites present during both log and stationary phase of growth. By combing this data with the genome scale metabolic model of yeast, we are able to follow the extracellular dynamics of metabolic usage and production as they enter log-phase, and transition into stationary phase. Furthermore, by identifying changes in intracellular metabolite concentrations, we are able to identify which metabolic pathways must change as the yeast enter stationary phase.

Our results provide some some suggestions metabolic mechanisms of dominance, which allow for the development of specific hypothesis to test. Answering these questions might allow for the engineering solutions for enhanced dormancy in plants, animals or humans, as well the development of anti-dormancy agents to target deleterious organisms such disease causing bacteria.

This work was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

163

Engineering Specialized Metabolism in a Single Cell Type

Anna Berim,[1] Adam Schmidt,[1] B. Markus Lange,[1] Michael Scanlon,[2] Carol A. Soderlund,[2] and **David R. Gang**[1]* (gangd@wsu.edu)

[1]Washington State University, Pullman; [2]Cornell University, Ithaca, New York; and [3]The University of Arizona, Tucson

Project Goals: The long term goal is to significantly increase our understanding of mechanisms that regulate and control specialized metabolism in single cell types of plants and to use that understanding to rationally engineer plants to produce desired metabolite profiles. The glandular trichome secretory cell is our model in vivo cell system. This project is divided into two focus areas. The first focus area involves development and testing of a model of metabolic partitioning between the MEP/terpenoid and shikimate/phenylpropanoid pathways in glandular trichome secretory cells. This model will be used to predict regulatory and control points in the large metabolic network, which then could be targets for modification to achieve desired alteration of metabolism in this single cell type. The second focus area will be the evaluation of the roles of specific protein modifications and of changes in their expression on carbon partitioning and metabolic flux in real world scenarios.

Mathematical Model Development and Data Collection
A number of advances were made to a previously generated kinetic mathematical model of peppermint essential oil biosynthesis. In particular, the Lange laboratory was able to demonstrate that the density and size distribution of glandular trichomes on peppermint leaves was the most important factor determining essential oil yield in various genotypes (wild-type and several transgenic plants) and under different adverse environmental conditions[1]. We considered using an analogous approach to model essential oil biosynthesis in basil, but there are critical gaps in our understanding of the physiology of trichome development and the organization of the terpene/phenylpropanoid pathways. Moreover, results over the course of the past year have supported the contention that significant amounts of non-volatile (ergo non-essential oil) metabolites accumulate in the glandular trichomes of basil. Based on these data, we are assembling a reaction graph representation of the pathways relevant to essential oil formation in sweet basil. We are using a Flux Balance Analysis (FBA) approach with one main difference (and key innovation) compared to existing FBA studies: we do not assume that biomass production is geared toward optimal growth (which works well in microorganisms but there are many problems with more complex organisms) but instead we are focusing our modeling on glandular trichomes. We are assuming the uptake of a transport sugar (stachyose/sucrose) and the maximizing of the production of essential oil and its removal from the network (accumulation in storage cavity of glandular trichome) as the objective function. These modeling activities will provide us with essential information regarding the flux distribution across multiple pathways in glandular trichomes. In addition, this approach does not require as much detailed kinetics data for the network elements as a kinetic model would require.

Advances in metabolite network determination
Work on characterizing the enzymes involved in production of the large array of methoxylated flavones in basil and peppermint has moved ahead very well. Because flavones, mostly highly methoxylated, constitute a considerable carbon sink both in basil and peppermint, which potentially competes with the isoprenoid pathway for carbon allocation, and the mechanisms of A-ring modifications have been barely studied hitherto, we have worked to characterize the late metabolic steps leading to the formation of these compounds. This information will be required for development and refinement of the model of metabolism in the secretory cells of basil. The functions of a number of enzymes involved in the formation of these compounds have been determined from basil and peppermint. For example, recombinant FOMTs display distinct substrate preferences and product specificities that can account for most detected 7-/6-/4'-methylated, 8-unsubstituted flavones in sweet basil. Apparent K_M values in the low micromolar range and specific gene expression profiles support the involvement of specific FOMTs in the biosynthesis of specific flavones in the different sweet basil lines. Structure homology modeling suggested the involvement of several amino acid residues in defining the proteins' stringent regioselectivities. The roles of these individual residues were confirmed by site-directed mutagenesis. A parallel study of flavone A-ring hydroxylases allowed us to delineate the network from apigenin to salvigenin, gardenin B and nevadensin, the major polymethoxylated flavones that accumulate in sweet basil.

Laser-microdissection of developing glandular trichome secretory cells for functional analysis and model refinement
The goal of this part of the project is to measure gene expression in glandular trichome secretory cells at specific developmental stages. This information is important in order to assign functions to the trichomes and not just leaf cells in general, and also for development and testing of the later mathematical models. We are using the technique of laser-microdissection coupled with RNA-sequencing to isolate specific cell types from sectioned tissue and to profile and quantify transcripts in these cells. Transcript profiles are being compared in peppermint and four basil chemotypes at four defined developmental stages from trichome initiation to fully expanded trichomes.

Reference

1. Rios-Estepa et al. (2010) Mathematical modeling-guided evaluation of biochemical, developmental, environmental and genotypic determinants of essential oil composition and yield in peppermint leaves. Plant Physiol. 152, 2105-2119

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

164

Characterization of Metabolites from Phototrophic Microbial Mats in Biogeochemically Unique Environments

Young-Mo Kim[1]* (young-mo.kim@pnnl.gov), Stephen R. Lindemann,[1] David M. Ward,[2] Sergey Stolyar,[1] Thomas O. Metz,[1] **Margaret F. Romine**,[1] and **James K. Fredrickson**[1]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash. and [2]Department of Land Resources and Environmental Science, Montana State University, Bozeman

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area is to develop a predictive, genome-enabled understanding of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. To achieve this goal, research is being conducted on three different model systems: iron- and sulfur-oxidizing communities associated with hyperthermophilic acidophilic springs at Yellowstone National Park (YNP) and phototrophic mat communities formed in high temperature springs at YNP or in a hyper-saline lake (Hot Lake) found in northern Washington State. As part of this effort, we are analyzing intracellular and extracellular metabolites to identify those that are secreted by microbes present in these phototrophic mat systems and to monitor their exchange among different community members.

The phototrophic microbial mats in YNP hot springs have been well-studied. However, the fundamental relationships between the primary producers (autotrophs) and consumers (heterotrophs) in these mats are not fully understood. Oxygenic phototrophic cyanobacteria, such as *Synechococcus* species, produce organic molecules by fixing inorganic carbon via photosynthesis or by fermenting stored molecules at night. For example, previous research has revealed that glycolate is produced during the day, while acetate is produced from the fermentation process at night. In turn, anoxygenic phototrophic bacteria, such as *Chloroflexus* and *Roseiflexus* species, and other heterotrophic organisms utilize these primary molecules for their carbon and energy needs. In this system, the metabolic processes of different community members contribute to the accumulation or consumption of storage molecules such as polyhydroxyalkanoic acids (PHAs) and wax esters for carbon storage and cyanophycin for nitrogen, which are thought to undergo diel cycling in *Roseiflexus* and *Chloroflexus*. PHAs are also possibly precursors for the synthesis of branched amino acids in certain bacterial species which have no enzymes to facilitate these processes. Using gas chromatography-mass spectrometry (GC-MS)-based metabolomics analyses, we have identified several key metabolites in this system. Some of these metabolites fluctuate during a diel cycle and may be exchanged between members of hot springs microbial mat communities. We will present results of metabolomics analyses of the mats, as well as the results of spent media analyses from the lab culture of *Thermosynechococcus* and *Chloroflexus* species which have been isolated from these systems.

The hyper-saline and ecologically unique Hot Lake in north-central Washington state has not been well-investigated by the scientific community. Recent studies have shown that several autotrophic and heterotrophic microorganisms comprise mats within this extreme environment, which contains almost 1 M magnesium sulfate and 0.5 M sodium chloride. Since the location of this lake is geographically isolated from any river or stream, organic carbon and nitrogen sources are not able to enter from outside sources. Therefore the majority of water input in this lake comes from precipitation, causing the lake to undergo fluctuations in salinity. The phototrophic mats in this system represent good models of energy and carbon fluxes. Importantly, the high levels of salts in this system present a challenge to the analysis of organic metabolites, which usually have similar size and polarity as those of salts. For example, high concentrations of salts may reduce extraction and derivatization efficiencies during GC-MS and generate excessive adduct ions in liquid chromatography-mass spectrometry (LC-MS) analyses. Therefore, we are developing and optimizing protocols to remove these salts to improve our measurements. We will present our on-going mass spectrometry-based metabolomics analysis of mat samples and secreted metabolites from cultures of autotrophic and heterotrophic organisms isolated from the Hot Lake community.

Results obtained from the above mentioned studies will be incorporated with data from transcriptomics and proteomics experiments and used for modeling the metabolism of the microbial mat communities.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Metabolite measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute through Contract No. DE-AC05-76RLO 1830.

165

Carbon Flow in Lithotrophic Acid Hot Springs Microbial Communities, Yellowstone National Park

Helen Kreuzer[1]* (Helen.kreuzer@pnnl.gov), James Moran,[1] Ryan Jennings,[2] Laura Whitmore,[2] Mark Kozubal,[2] Bill Inskeep,[2] and **Jim Fredrickson**[1]

[1]Pacific Northwest National Laboratory, Richland, Wash. and [2]Montana State University, Bozeman

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area is to develop a predictive, genome-enabled understanding

of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. To achieve this goal, research is being conducted on both natural mat communities and defined co-cultures designed to test hypotheses regarding processes naturally occurring in these communities. As part of this effort, we are elucidating carbon flow in the lithotrophic microbial communities in acid thermal springs in Yellowstone National Park, focusing particularly on the roles of autotrophy and heterotrophy in the communities.

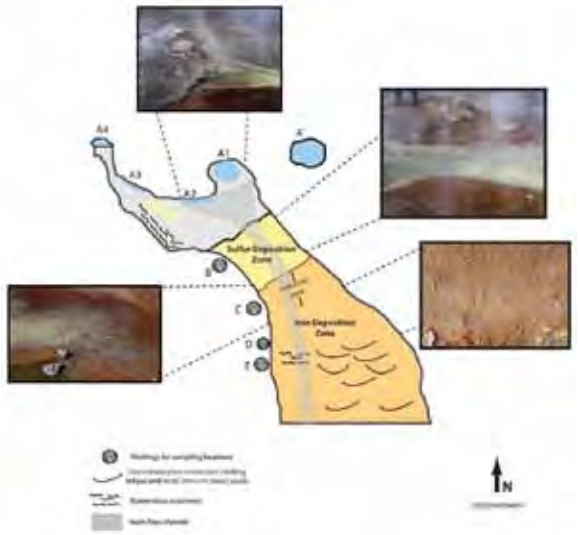


Figure 1. Beowulf spring in Norris Geyser Basin

Norris Geyser Basin is home to several acidic (pH ~3), sulfidic, hydrothermal (source temperature >65 ˚C) springs, of which Beowulf, Grendel, and OSP Springs are examples. Microbial communities inhabit these springs that are based on chemolithotrophy as evidenced by the oxidation of sulfide and ferrous Fe and subsequent deposition of oxidation products. What remains unclear, however, is how carbon is initially introduced into the systems; whether by autotrophy, heterotrophy, or a combination route. We are exploring carbon cycling in chemotrophic springs in the Norris Geyser Basin to better understand both the initial carbon entry into the system and how fixed C is exchanged in the microbial community. Our immediate goals include (1) determining whether autotrophy plays a major role in carbon incorporation, focusing initially on *Metallosphaera yellowstonensis*, a community member whose genome includes genes from the 3-hydroxypropionate 4-hydroxybutyrate carbon fixation pathway, and (2) compiling a stable isotope inventory of an example spring (Beowulf, Figure 1) and the surrounding landscape to help identify additional carbon sources.

By understanding how carbon is incorporated into and exchanged among microbial community members, we will gain insights into how microbial communities adapt to these extreme (high temperature, acidity, high sulfide and low biomass) environments.

Visual evidence reveals the two dominant metabolisms, shown in main Beowulf flow channel (Figure 1). Sulfide oxidation converts dissolved sulfide into elemental sulfur in zone B of the spring. The orange precipitates downstream of the sulfur deposition zone result from iron oxidizer activity.

M. yellowstonensis was isolated from Norris Geyser Basin and represents up to 20% of the total microbial nucleic acid associated with the iron oxide mats. *M. yellowstonensis* contains genes consistent with its presumed iron oxidizing metabolism.[1] Further, *M. yellowstonensis* has genes associated with the 3-hydroxypropionate, 4-hydroxybutyrate carbon fixation pathway. We sought to determine whether *M. yellowstonensis* actually does fix CO_2, and whether autotrophy occurs in the spring communities.

Pure cultures of *M. yellowstonensis* were incubated in mineral medium with or without yeast extract, and with un-labeled or ^{13}C-labeled CO_2 in the headspace. Ground pyrite was supplied as the electron donor, and O_2 was the electron acceptor. Cultures were incubated approximately two weeks, and the carbon isotope ratio of the acid-washed biomass was measured by isotope ratio mass spectrometry. When $^{13}CO_2$ was present, the isotope ratio of the biomass contained significantly more ^{13}C than when unlabeled CO_2 was present, demonstrating that *M. yellowstonensis* fixes CO_2. Much less ^{13}C was incorporated when yeast extract was present, suggesting that *M. yellowstonensis* grows heterotrophically in the presence of suitable organic substrates. Initial estimates of the CO_2 –> biomass fractionation factor (~3.5 ‰) are consistent with the hydroxybutyrate fixation pathway[2] and very similar to that measured for *M. sedula* (3.1 ‰).[3]

Samples of iron-oxidizing microbial mat from two similar Norris Geyser Basin springs, Grendel and OSP, were excised and placed into bottles containing pyrite and the mineral medium used in the laboratory experiments. Unlabeled CO_2 or $^{13}CO_2$ was injected into the headspace, and the cultures incubated fourteen days at 65 -70 ˚C. Following incubation, the C isotope ratio of the acid-washed mat material was measured by isotope ratio mass spectrometry. Data clearly demonstrate $^{13}CO_2$ fixation by the microbial communities in both springs.

Our data thus far indicate that there may be additional carbon inputs to the iron mat community. The bulk carbon isotope ratios we have measured in the iron oxide mats range from ~=18 to ~-21 ‰. Based on the approximate fractionation factor measured in the laboratory, autotrophic fixation of dissolved inorganic carbon by *M. yellowstonensis* would yield fixed carbon with a predicted isotope ratio of ~-7.5 ‰. A simple metabolic model for the iron mats, in which carbon is fixed by *M. yellowstonensis* and then passed through a heterotrophic food chain (which typically results in little isotopic fractionation) is not consistent with the measured bulk carbon isotope ratio in iron oxide mat samples. Alternative hypotheses to explain the bulk carbon isotope ratio of the mats include (1) the presence of significant landscape carbon (e.g., plant detritis) in the mats, or (2) direct heterotrophy of landscape carbon by members of the iron mat community. The average carbon isotope ratio of organic material in the landscape (plant material, animal dung) is ~ -26 to

-29 ‰, and that of dissolved organic carbon in Beowulf Spring is ~ -22 to -23 ‰. A mixture of passively present landscape carbon, or its heterotrophic uptake, in combination with carbon fixation by *M. yellowstonensis*, could explain the observed bulk isotope ratios. In the coming year, we will be working to associate carbon substrates with different community members by measuring isotope ratios of the different phylogenetic groups, and to further elucidate the flow of carbon through the iron mat community.

References

1. Kozubal, M.., et al, 2011. *Appl. Environ. Microbiol.* 77(5):1844.
2. Berg, I., et al., 2010. *Nature Rev. Microbiol*. 8:447.
3. House, C., et al., 2003. *Org Geochem*, 2003 34:345.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

166

Metabolic and Community Modeling of Phototroph-Heterotroph Community Interactions

William Cannon[1]* (william.cannon@pnl.gov), Gretta Serres,[2] Sebastian Jaramillo,[1] and Don Bryant[3,4]

[1]Computational Biology and Bioinformatics Group, Pacific Northwest National Laboratory, Richland, Wash.; [2]Marine Biological Laboratory, Woods Hole, Mass.; [3]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park; and [4]Department of Chemistry and Biochemistry, Montana State University, Bozeman

Project Goals: Understanding interactions between microbes in the environment is a daunting task. Because of the complexity of such interactions, it is necessary to use computational models to track the data and proposed interactions, and to develop intuition, hypotheses and predictions. With this in mind, the goals of the modeling component of the PNNL FSFA are two-fold. First, in the short-term we will evaluate the existing metabolic model of *Synechococcus* sp. PCC 7002 by incorporating metabolomic and proteomic data. We will initially utilize the model to evaluate whether the newly discovered complete but alternative cyanobacterial TCA cycle maximizes the utilization of 2-oxoglutarate for production of biosynthetic precursors for growth, and to evaluate the growth and production of extracellular metabolites. Second, in the long term, we will more generally identify the key metabolic interactions between the autotrophic and heterotrophic populations of Yellowstone hot spring mats and consolidate the quantitative data describing these interactions within a context of extended phototroph–heterotroph model, including metabolite predictions.

Building useful models requires several tasks. First, initial models are constructed based on genome sequences. These initial models only tell us about metabolic potential, rather than actual metabolism because genome annotation is not complete and regulatory information is missing. Enzymes and transporters that carry out reactions of interest may not be expressed under all conditions, predicted metabolites may not actually be produced, and more frequently unpredicted metabolites are observed. Thus a second critical aspect to building useful models is to have data analysis tools that analyze the data thoroughly. Multiple data types—proteomics, metabolomics and transcriptomics to name a few—may then be interpreted with respect to the model. Based on the data, the model is adjusted. This is the process frequently referred to as data integration. Finally, the model is curated to the point at which simulations can be carried out, the output of which can be directly compared to experimental data.

To enable process of model building, the PNNL SFA has developed 11 pathway genome databases of phototrophic and heterotrophic species being studied on this project. These databases are linked to external pathway databases of the same organisms for comparative studies. Using this platform, experimental data can be interpreted and used to refine the model.

Using this approach, we evaluated the protein complement of *Synechococcus* spp. 7002 grown under conditions of phototrophic growth under atmospheric CO_2 concentrations. Not surprisingly, pathways related to photosynthesis and carbon fixation were identified most frequently. However, initial analysis implied that a specific ICT family CO_2 transporter, SYNPCC7002_A0690, was not expressed, implying that either the annotation was incorrect or an alternative transporter is used. Since transporters set the boundary conditions that determine the energetic feasibility of metabolic processes, such adjustments to the metabolic model are critical. However, a new, highly accurate computational method that we applied found that the ICT transporter SYNPCC7002_A0690 was not only expressed but appears to be one of the most highly expressed proteins in the cell [1].

Pathway databases also serve as sources of hypotheses. Using pathway databases, we have begun to investigate pathways that have been inferred to be incomplete by computational methods. Investigation into the incomplete TCA cycle of cyanobacteria has demonstrated that, in fact, the cycle is complete and uses an alternative enzyme for the conversion of 2-oxoglutarate to succinate [2]. A principle difference between the standard pathway and the newly discovered alternative pathway is that the formation of succinate is no longer coupled to the highly favorable reaction involving coenzyme-A. Due to the removal of this highly favorable step, we hypothesize that the alternative pathway leads to an increased steady-state level of 2-oxoglutarate. This could be significant in that 2-oxoglutarate is the precusor used for assimilation of nitrogen, and the *Synechococcus* species in the Yellowstone mat system have been observed to dramatically increase nitrogen fixation during a brief 2-hour period in

the morning. This burst of activity may serve as a nitrogen source for other members of the microbial community, and yet it is not clear how such a dramatic increase in nitrogen fixation can occur in such a short time period. Both laboratory and simulation based tests of these hypotheses are being conducted.

The simulation model starts with the framework of the previously determined genome-scale and flux-based model of *Synechococcus* sp. PCC 7002. We will model the TCA cycle in kinetic detail by incorporating aspects of the kinetic model by Singh and Ghosh. This will allow us to predict 2-oxoglutarate levels for both the standard and alternative TCA cycles under multiple conditions.

References

1. W. R. Cannon, M. M. Rawlins, D. J. Baxter, S. J. Callister, M. S. Lipton, and D. A. Bryant, "Large Improvements in MS/MS-Based Peptide Identification Rates using a Hybrid Analysis," *J Proteome Res,* Mar 30 2011.
2. S. Zhang and D. A. Bryant, "The Tricarboxylic Acid Cycle in Cyanobacteria," *Science,* vol. 334, pp. 1551-1553,2011. http://www.sciencemag.org/content/334/6062/1551.abstract

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

167

Metabolic Coupling and Carbon Flux Regulation in Phototroph-Heterotroph Associations

Alex S. Beliaev* (alex.beliaev@pnnl.gov), Thomas C. Squier* (thomas.squier@pnnl.gov), Diana J. Bigelow, Grigoriy E. Pinchuk, Sergey M. Stolyar, Baowei Chen, Na Fu, Yijia Xiong, E.A. Hill, and **Jim K. Fredrickson**

Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The overarching goal of the PNNL Foundational SFA is to understand the collective energy, carbon, and nutrient processing in phototrophic microbial communities that contributes to their stability and efficient utilization of resources. As part of the FSFA cross-cutting science and technology theme, this research component elucidates pathways and regulatory mechanisms that govern metabolic exchange and interactions between organisms in phototrophic associations and contribute to their stability and efficient utilization of resources.

Scientific Concept: Secretion of low-molecular weight metabolites and biopolymers drives interactions between primary producers (autotrophs) and consumers (heterotrophs) and, at the organism level, is a function of environmental parameters such as light intensity, O_2 levels, and nutrient concentrations. Under photosynthetic conditions, O_2 is one of the key factors affecting carbon metabolism, growth, and survival of microorganisms. We hypothesize that photoautotroph-heterotroph associations have developed mechanisms to spatially and transiently decrease O_2 concentrations thereby lowering oxidative stress. Under dark conditions, the phototrophs ferment glycogen providing the heterotrophs with electron donor/carbon source to carry out anaerobic respiration which affects solubility of critical micronutrients such as Fe and Mn, thus making these elements more accessible for phototrophs. At a sub-cellular level, the interactions are mediated by multi-level regulatory networks that include redox-active proteins which sense environmental change through specific post-translational modifications and act as key control points to optimize metabolic efficiencies at both organism and community levels. To that end, we seek to identify regulatory proteins responsive to environmental change throughout the diel cycle, which includes redox-dependent disulfide bond formation and oxidative modifications to Met and Tyr (oxidative modifications are reversible through the action of repair enzymes). We hypothesize that light-responsive disulfide bond formation acts to control the functions of key proteins that regulate metabolic fluxes. Such adaptive regulation is suggested to diminish formation of reactive oxygen species (ROS), reduce protein oxidative damage, and enhance metabolic efficiencies. *In vivo* examples have been identified involving an inverse relationship between redox-dependent cystine formation and the abundance of site-specific oxidative modifications. As these post-translational modifications are reversible, they have the potential to rapidly reprogram metabolism through the control of specific branch points within metabolic pathways.

Our current line of research focuses on gaining an understanding of the pathways and regulatory mechanisms associated with the collective energy, carbon, and nutrient processing in axenic and mixed phototroph-heterotroph cultures. Systems under investigation include *Synechococcus* sp. PCC 7002 – *Shewanella putrefaciens* W3-18-1 co-culture, individual organisms (*e.g., Synechococcus* spp., *Thermosynechococcus* sp., *Roseiflexus castenholzii*, *Chloroflexus* spp.), and consortia derived from, hot spring communities (e.g., *Thermosynechococcus* sp. N55-*Roseiflexus castenholzii*). This research also utilizes advanced controlled cultivation capabilities to enable investigation of redox-dependent control mechanisms.

Baseline growth parameters and the presence of regulated metabolic coupling between *Synechococcus* and *Shewanella* spp. have been identified, as both organisms successfully grow as a binary co-culture, but not as separate monocultures, using either lactate or CO_2 as the sole source of carbon. During growth with lactate, O_2-dependent oxidation of lactate mediated by W3-18-1 produced sufficient amounts of CO_2 to maintain photosynthetic growth of 7002, which in turn, generates the necessary O_2 for growth of W3-18-1. These results establish that there is a tight metabolic coupling between the phototrophic and heterotrophic organisms using light as the only source of energy and lactate as the sole source of carbon. Growth of the co-culture under these conditions results in acetate accumulation by *Shewanella* W3-18-1 thus revealing an imbalance is caused

by differences in growth rates where the O_2 consumption by the heterotroph exceeds the O_2 evolution rates by the cyanobacterium. When grown on CO_2 as the sole source of carbon, the metabolic coupling between 7002 and W3-18 is established through secretion of dissolved organic carbon compounds by the photoautotroph. NMR-based techniques identified formate, lactate, and acetate secreted by *Synechococcus* 7002 which may serve as the primary carbon and energy sources for *Shewanella* W3-18-1. Secretion of extracellular metabolites and/or biopolymers also appears to be common for some thermophilic cyanobacterial species, e.g., *Thermosynechococcus* sp. In hot spring phototrophic (HSP) mat communities major shifts in energy metabolism occur in response to increases in O_2 involving: *i*) alterations in metabolic exchange between community members through the release of acetate, and *ii*) diminished rates of carbon fixation due to photorespiration that may be associated with an uncoupling of RuBisCo and futile cycling involving glycolate production. These results indicate that enhanced metabolic efficiencies in communities of phototrophs and heterotrophs may arise, in part, due to increased efficiencies of carbon fixation as a result of reductions in photorespiration due to decrease in O_2 levels.

Additional cellular responses that impact community stability were found using chemical probes; they involve regulatory mechanisms that shift metabolic flux to minimize the oxidative damage to cellular proteins in a manner that will enhance metabolic efficiencies. Large decreases in the overall levels of protein oxidation are observed for cultures of 7002 grown in the presence of the W3-18-1, where the majority of oxidatively sensitive proteins in W3-18-1 are protected from oxidative modification in the co-culture. Reductions in oxidative stress are apparent despite substantially higher O_2 present in the co-culture (160% dissolved air saturation) in comparison to axenic cultures (44% dissolved air saturation) (both cultures are grown in caged photobioreactors using white light intensities of 240 μM photons m^{-2} s). These results support the hypothesis that opportunistic interactions between heterotrophic (*Shewanella*) and photosynthetic (*Synechococcus*) organisms permit metabolic coupling to enhance energy efficiencies and community stability. These measurements are consistent with prior indications that axenic isolates of the *Synechococcus* isolated from the HSP mats have a substantially enhanced sensitivity to light-induced oxidative stress in comparison to the natural mat community, indicating an importance of metabolic coupling between community members in the mat that promote stress resistance.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

168

Microbial Diveristy and Biogeochemical Function of the Phototrophic Microbial Mats of Epsomitic Hot Lake, Washington

Stephen R. Lindemann* (stephen.lindemann@pnnl.gov), James J. Moran* (james.moran@pnnl.gov), Alice Dohnalkova, David W. Kennedy, Sergey M. Stolyar, Paul D. Majors, H. Steven Wiley, Allan E. Konopka, and **James K. Fredrickson**

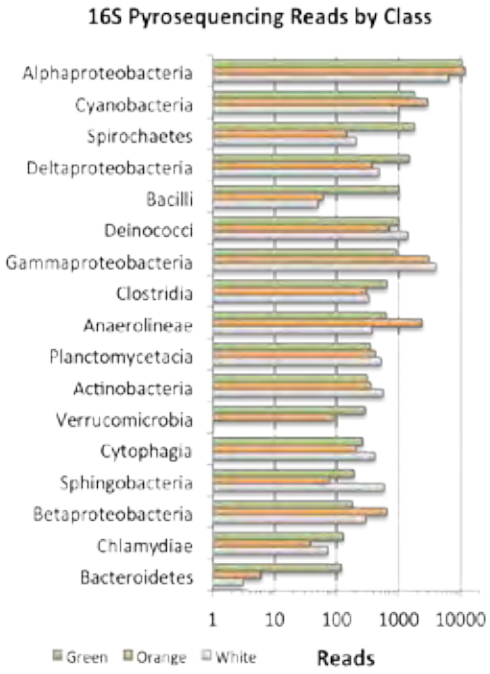
Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Determine the phylogenetic and functional diversity of members of a hypersaline microbial mat using (meta)genomics of whole mat, enrichment cultures, and sorted single cells; probe the dynamics of community function and metabolic interactions in a hypersaline microbial mat in response to changes in salinity, temperature, and photon flux using metatranscriptomics and stable isotope probing; ultimately, to generate a whole-mat model of energy and element cycling in the mat within its ecological context, especially with respect to carbon balance.

Hot Lake is a meromictic, hypersaline lake that occupies a small glacial endorrheic basin in north-central Washington and contains a benthic phototrophic microbial mat. Being highly organized, metabolically interactive, self-sustaining communities, phototrophic microbial mats are natural models for the study of energy and element cycling between phototrophs and heterotrophs and the effect these interactions have upon biogeochemical processes, especially with respect to carbon cycling. As there are no water inputs into Hot Lake besides meteoric water falling within the small basin it occupies, primary production within the lake is thought to be the major source of fixed carbon; this simplifies analysis of the carbon and energy budget of the phototrophic mat and the mat's influence on its broader geochemical context. Carbonate minerals precipitate within the mat beneath the cyanobacterial stratum by an as yet-unknown mechanism. As the metabolic activity of both sulfate reducers and oxygenic phototrophs have been independently implicated as the key driver of biogenic carbonate precipitation within microbial mats in disparate environments, study of the interactions between Hot Lake mat phototrophs and heterotrophs will shed light on the biological mechanisms of carbonate deposition common to epsomitic environments. Furthermore, Hot Lake is a dynamic system; besides variability in temperature and photon flux throughout the annum, Hot Lake also experiences a seasonal cycle in salinity, varying between observed extremes in epilimnion water of ~200 mM $MgSO_4$ after snow melt to greater than 2 M at the close of the dry season. This cycle permits analysis of the stability in composition and function of the same community under diverse chemiosmotic conditions and elucidation

of the mechanisms that impart robustness. As changes in climate are likely to alter both the temperature and salinity of surface waters, the study of Hot Lake may also improve our prediction of the effects of climate change on microbial community function and biogeochemical cycling, as well as offer insight and bioprospecting for high-salt (especially epsomitic) industrial and bioenergy applications.

The first aim of this project is to understand the dynamics of community structure and function in response to the annual cycle of the major physicochemical parameters of Hot Lake, especially salinity, temperature, and photon flux. Initial analysis of community diversity by 16S rRNA pyrosequencing of mat obtained during near the seasonal maximum in photon flux revealed taxonomic units consistent with oxygenic and anoxygenic photosynthesis, sulfur oxidation and reduction, and halotolerance. Nearest-neighbor phylogeny inferred two main filamentous cyanobacteria, one of order *Oscillatoriales* and one of order *Nostocales* present within the mat. Halophilic purple sulfur bacteria of genus *Halochromatium* and *Thiohalocapsa* as well as purple non-sulfur bacteria such as *Roseobacter* and *Rhodovulum* compose the anoxygenic phototrophic guild. 16S sequences consistent with the presence of bacteria similar to the nitrate-reducing, obligately chemolithotrophic sulfur oxidizer *Thialkalivibrio nitratireducens* as well as the halophilic deltaproteobacterium *Desulfosalina* suggests active sulfur cycling within the mat. Samples were collected from three points along the depth/salinity gradient of the lake; while sequences from the major members of the expected major functional guilds were observed in all three samples, elevations in *Desulfosalina*, *Spirochaeta*, and *Bacteroidetes* in the region of the halocline suggest an increased role for anaerobic degradation of mat biomass in the transition from the oxic epilimnion to the anoxic hypolimnion. Seventeen unique strains have been isolated from the mat to date, including a previously uncultivated *Rhodobacteriaceae*, *Marinobacterium*, and *Salinibacterium* species. Both of the primary oxygenic phototrophs have been grown in unicyanobacterial consortia with mat heterotrophs with bicarbonate as the sole carbon source. These cultures are currently being analyzed by HR-NMR and LC-MS to determine the nature of the organic compounds that are being exchanged between the cyanobacterial phototrophs and associated heterotrophs. Furthermore, under a JGI Community Sequencing Program project, metagenomes of these co-cultures and whole mat will be generated and single cells from interesting taxa derived from whole mat will be isolated and sequenced. Paired samples will be utilized for metatranscriptomic analysis of *in situ* mat collected around the diel cycle. The resultant sequences, in concert with empirical data from naturally-derived simplified communities, will be used for predictive modeling of community membership and function, with an emphasis upon exchanged metabolites, in response to changes in physicochemical parameters.



The second major goal of this project is to generate a whole-mat model of energy and element cycling in the mat within its ecological context, especially with respect to carbon balance. Initial efforts towards this goal are focused on broad chemical classes; including dissolved organic carbon, sediment carbon, mat carbon, and surrounding landscape carbon. We are inventorying these pools to quantify where carbon is stored in the system and employing NMR and mass spectroscopy to identify key chemical species within each of these larger inventory pools. Stable isotope analysis is being used to link different pools and help identify carbon sources metabolized within the mat. Such surveys of total carbon and inventories of specific carbon compounds allow detection of connections between mat metabolic activity and the surrounding geochemical landscape. We are exploring

metabolites produced in representative cultures and enrichments from Hot Lake to identify important potential targets for use in future SIP experimentation. This groundwork will provide a foundation for *in situ* stable isotope probing experiments to occur in the coming field season. A solid understanding of the carbon landscape within and surrounding the mats coupled with the SIP experiments should illuminate potential key carbon currencies in the mat and provide clues to carbon exchange between species within the mat community.

We are also working to understand how the system geochemistry influences the mat community. We are using a simple geochemical model to predict supersaturated species in the lake over very large seasonal fluctuations in water temperatures, salinity, and redox conditions. Initial measurements of microgradients within *ex vivo* mats have been measured with dissolved oxygen, pH and redox microelectrodes over a variety of light regimes, which is being coupled to the predictive model to refine our understanding of what geochemically-important processes are occurring within microenvironments inside the mat.

The combined genomic, enrichment incubation, and geochemical data collected to date will be compiled to guide sample collection during the summer of 2012, culminating in generation of a comprehensive dataset permitting multifaceted analysis of the activity of the mat and interactions of its members around the diel cycle near the seasonal photic maximum.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Electron microscopy and NMR measurements were performed in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by OBER and located at PNNL.

169

Microbial Interactions in Hot Spring Phototrophic Mats and Isolates

David M. Ward[1]* (umbdw@montana.edu), Christian Klatt,[1] Donald A. Bryant,[1,2] Zhenfeng Liu,[2] and Sergey Stolyar[3]

[1]Montana State University, Bozeman; [2]The Pennsylvania State University, University Park; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: Microbial mats in alkaline siliceous hot springs of Yellowstone National Park, which are constructed by a variety of oxygenic and anoxygenic phototrophic microorganisms, are model systems from which to learn principles that underlie a predictive understanding of microbial community ecology. Our goal is to use multiple "omics" and cultivation approaches to understand microbial interactions that govern the capture and recycling of energy and resources in this system.

Analyses of assembled metagenomic scaffolds have identified eight major functional guilds inhabiting the photic portion of the community. These include the oxygenic phototrophic *Synechococcus* spp., five anoxygenic phototrophic guilds (*Roseiflexus* spp., *Chloroflexus* spp. and Anaerolineae-like Chloroflexi; "*Candidatus* Chloracidobacterium thermophilum" and "*Ca*. Thermochlorobacter aerophilum") and two aerobic heterotrophic populations. Detailed annotations of genes in these assemblies have permitted reconstruction of the metabolic potential within each guild and this has guided us in developing a model of how the members of different guilds might interact metabolically. For instance, consistent with previous results, *Synechococcus* spp. possess genes that should enable them to produce glycolate through photorespiration and to produce glycogen and break it down via fermentation. *Roseiflexus* spp., "*Ca*. C. thermophilum", " *Ca*. T. aerophilum" and members of heterotrophic guilds possess genes that should enable them to metabolize glycolate and/or acetate. Surprisingly, "*Ca*. C. thermophilum" and "*Ca*. T. aerophilum" are both missing the same genes involved in the synthesis of branched amino acids (though not the transaminase genes). This observation suggests that they may obtain relevant biosynthetic precursors from another mat organism(s). One hypothesis is that polyhydroxyalkanoic acids (PHAs) of *Roseiflexus/Chloroflexus* spp. which are produced from related compounds can be a source of branched amino acid precursors for above mentioned organisms. Additionally, *Roseiflexus* spp. appear to be deficient in the synthesis of riboflavin, suggesting their dependence on populations that can produce it (e.g., *Synechococcus* spp., "*Ca*. T. aerophilum"). Finally, both *Synechococcus* spp. and *Roseiflexus* spp. may have potential to fix N_2, possibly suggesting that other community members may be dependent on them for a supply of fixed nitrogen.

These metagenomic assemblies have also been useful as references for interpreting metatranscriptomics and metaproteomics databases. While transcription patterns in general confirmed previous observations, they also revealed new ways to think about the metabolisms of mat community members. For instance, it was formerly thought that *Roseiflexus* spp. were primarily photoheterotrophic during daytime, aside from a brief period of light, but anoxic conditions in early morning, when they conduct carbon fixation via the 3-hydroxypropionate pathway. It was thus surprising to find that genes in this pathway were expressed during the day. Detailed examination of expression patterns for genes involved in central carbon metabolism and storage polymer synthesis and degradation now lead us to hypothesize that *Roseiflexus* spp. conducts a photomixotrophic metabolism throughout the day. We hypothesize that to do so they produce PHAs (and possibly wax esters) at night, which they use during the day to maintain an adequate supply of organic intermediates and reductant. If true, this could have important implications on the timing of transfer of intermediates derived from PHAs to other members of community.

These hypotheses are being tested in several ways. First, to demonstrate that *Roseiflexus/Chloroflexus* spp. incorporate CO_2 in the low-light, anoxic morning period and also in high-light oxic period, we investigated $^{13}CO_2$ incorporation

into proteins of these organisms *in situ.* In a similar manner, ^{13}C-acetate has been used to assess the degree to which *Roseiflexus* spp., "*Ca*. C. thermophilum," and "*Ca*. T. aerophilum" may compete for this intermediate. Second, metabolomics analyses are being used to evaluate the hypothesis that PHAs (and possibly wax esters) undergo diel rhythms of synthesis and degradation in *Roseiflexus/Chloroflexus* spp. Third, we are developing experimental co-cultures consisting of cyanobacteria and Chloroflexi isolated from hot spring cyanobacterial mats, namely, *Thermosynechococcus sp. NAK55* (whose genome we obtained), *Roseiflexus castenholzii* and *Chloroflexus aurantiacus* J-10-fl. We found that *Thermosynechococcus* excreted a variety of metabolites such as formic, lactic, succinic, oxalic and isocaproic acids, dihydroxyacetone, sarcosine and glyceraldehyde, which have not yet been considered in our model of metabolic interactions inferred from metagenomic data. *Choloroflexus* J-10-fl was tested for its ability to utilize these metabolites and some of them were found to support growth. To investigate these potential interactions further, we are studying growth of the cyanobacterium in co-cultures with Chloroflexi isolates. Metabolites present in the growth medium and gene expression in the members of the co-cultures are being determined under light and dark conditions in two experimental settings: i) slowly mixed planktonic culture and ii) mat forming, benthic cultures growing on the surface of submerged agar. These investigations are enhancing our understanding of metabolic interactions among members of major guilds in natural hot spring cyanobacterial mats and providing new experimental approaches to evaluate the interactions.

This research is part of the PNNL Foundational Scientific Focus Area: Biological Systems Interactions.

170

Predicting Autotroph—Heterotroph Metabolic Interactions from Genome Sequence

Margrethe (Gretta) Serres[1]* (mserres@mbl.edu), Elizabeth Landorf,[2] Frank Collart,[2] Sebastian Jaramillo Riveri,[3] and **Margie Romine**[3]

[1]Marine Biological Laboratory, Woods Hole, Mass.; [2]Argonne National Laboratory, Lemont, Ill.; and [3]Pacific Northwest National Laboratory, Richland, Wash.

Project Goals: The PNNL Foundational Scientific Focus Area, Biological Systems Interactions, investigates microbial interactions to understand how microbial communities work collectively to carry out complex biogeochemical processes. We seek to understand how the exchange of metabolites between organisms contributes to the stability of microbial communities. A longer-term goal is to develop predictive capabilities for the response of microbial communities to environmental change. A model system involving *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 was selected to develop and test predictions of molecular exchange processes occurring during co-culture growth. To examine the molecular basis for this phototroph:heterotroph interaction we are applying comparative sequence analysis and experimental approaches to define and describe features in the genomes associated with their interactions and thereby identify functional traits important to microbial community structure. These investigations support lab-based investigations examining growth, physiology, metabolite exchange, and gene expression patterns in axenic and co-cultures.

The two model organisms chosen for our study, *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1**,** have been co-cultured without the supplement of nutrients supporting that their metabolisms can be coupled. *Synechococcus* sp. PCC 7002 is an oxygenic photoautotroph also capable of fermentative metabolism that produces different metabolic end products under variable environmental (i.e. light) conditions. *Shewanella* sp. W3-18-1 is a facultative anaerobe capable of utilizing a variety of carbon and energy sources.

Automated genome annotations of *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 were curated to improve the specificity of functional assignments and to identify key knowledge gaps (e.g. missing steps in metabolic pathways, missing transport reactions) that require further investigation via computational and wet-lab methods. Gene models for both organisms have been adjusted using proteome data to validate protein starts and sub-cellular locations. Function predictions have included the use of experimental data from well-studied model genomes, such as *Synechocystis* sp. PCC 6803 (*Synechococcus*) and *S. oneidensis* MR-1 (*Shewanella*) as well as orthologs from phylogenetically related organisms. Pathway genome databases have been constructed for the *Synechococcus* and *Shewanella* pair. Regulatory interactions inferred in RegPrecise are included in the curation.

Comparative analysis of the *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 genomes enables identification of the metabolic pathways that are unique to each organism. These pathways and their corresponding metabolites are a source of potentially exchanged nutrients that can contribute to the interspecies metabolic interactions. We also include pathways associated with compounds known to be present in autotrophic:heterotrophic consortia (i.e. glycolate, propionate, lactate, acetate). Based on our pathway predictions we are cataloging potential interaction points for the *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 pair. To establish the molecular connection between these cellular metabolic pathways and exchange nutrients, we generated a profile for the transport capabilities of these organisms. Inferred functional attributes were identified by sequence comparisons using Blast and reconciled with predictions from the Transporter Classification Database. To validate inferred transport functions, we examine the ligand binding preference for the ABC transporters. This family of transporters is ubiquitous in bacterial systems and provides import and export capability for a wide spectrum of ligands. A typical ABC importer generally consists of a periplasmic solute-binding protein, two integral membrane subunits, and two cytoplasmic ATPases. It is the solute binding proteins (SBPs) of the ABC transport complex that recognize and bind specific substrates in the cell wall and transfer them to

the membrane subunits, and are therefore responsible for the uptake of ligands from the environment. We screened and characterized a set of transporter proteins from *Synechococcus* sp. PCC 7002 and *Shewanella* sp. W3-18-1 using a fluorescence thermal shift assay (FTS) and isothermal calorimetry. These functional screens, that included metals, small ions, mono- and oligosaccharides, peptides, amino acids, polyamines and vitamins, generated specific binding ligand assignments for approximately 60% of the purified and screened proteins. The experimental studies provide new functional information for these transport complexes and also validate many of the assignments derived from comparative analysis. These transport capabilities support the identification and characterization of metabolic and regulatory pathways for these organisms and provide a basis for experimental validation of the potential exchange nutrients that contribute to the interspecies metabolic interactions.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

171

Chemical Profiling of Group B Vitamin Transport and Protein Interactions in Microbes and Microbial Communities

Aaron T. Wright[1]* (Aaron.Wright@pnnl.gov), Phillip K. Koech,[1] Andrei L. Osterman,[2] Dmitry A. Rodionov,[2] and **Margie Romine**[1]

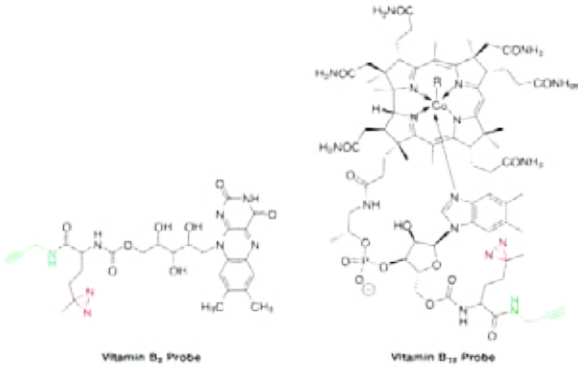
[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash. and [2]Sanford-Burnham Medical Research Institute, La Jolla, Calif.

Project Goals: The objective of the Pacific Northwest National Laboratory Foundational Scientific Focus Area is to develop a predictive, genome-enabled understanding of how microbial interactions impart stability, robustness, and functional efficiency to microbial communities. To achieve this goal, research is being conducted on both natural mat communities and defined co-cultures designed to test hypotheses regarding processes naturally occurring in these communities. As part of this effort, we have initiated a new line of investigation focused on identifying the transporters and sensors that mediate the exchange of materials between autotrophs and heterotrophs present in these communities. To complement ongoing genetic and *in situ* approaches aimed at testing substrate specificity, we propose to utilize chemical protein profiling to test specific substrates that are hypothesized to be exchanged in these communities, starting with B-type vitamins.

Transporters and sensor proteins are the cellular interface with the environment and the conduit for exchange of molecules and transfer of information in the ecosystem. However, their specificity cannot be readily accurately predicted from genome sequence information alone, making it difficult to predict the identity and nature of interacting microbial partners without further experimental investigation.

Chemical protein profiling employs synthesized probes that consist of three chemical elements: a binding group that biases the probe toward a protein family and/or is a substrate-mimic; a diazirine moiety for irreversibly labeling proteins; and a reporter tag for detection and isolation of probe-labeled proteins from biological systems. Probe synthesis will exploit the multimodal bio-compatible click chemistry (CC) reaction to create "tag-free" probes for profiling proteins in living systems. Probe-labeled proteins will be visualized by addition of a complementary azide-modified reporter tag using the CC cycloaddition reaction. Common reporter groups include fluorescent tags such as azido-tetramethylrhodamine for gel (SDS-PAGE)-based analysis, or an azido-biotin tag for enrichment and liquid chromatography-mass spectrometric analysis (LC-MS). The method permits a variety of characterization techniques including, but not limited to: live-cell fluorescent imaging, fluorescent SDS-PAGE, fluorescence polarimetry, FACS sorting, and LC-MS analysis. Anticipated applications include evaluation of molecular function, cellular distribution or uptake, characterization of probe-labeled proteins by MS-based proteomic analysis, and real-time monitoring of probe uptake and protein binding. We will integrate our experimentally identified metabolite-protein interactions with computational efforts to characterize novel sequence variants of transporter and sensor proteins.

As one of our first objectives we intend to characterize transporters and sensors that modulate group B vitamin exchange in microbial co-cultures and in natural communities. These vitamins are required, in relatively limited amounts, as intermediates/precursors in the biogenesis of key cofactors in central metabolism. Subsystems analysis revealed that the presence of *de novo* and/or salvage pathways is highly variable among representatives of microbial taxa in the natural communities under investigation and thus vitamin exchange is a likely driver of opportunistic interactions. Our initial analyses will involve the use of vitamin B_2 and vitamin B_{12} chemical probes to define group B vitamin transport/sensor interactions in a *Synechococcus* sp. PCC 7002 and *Shewanella putrefaciens W3-18-1* co-culture. Comparative analysis of subsystems and regulons suggest that *S. putrefaciens W3-18-1* is able to synthesize B_2 and B_{12} while *Synechococcus* sp. PCC 7002 can only synthesize B_2. The proteins required for uptake of vitamins B_2 and B_{12} have been predicted in each organism, however those involved in its export (if any exist) by *Shewanella* are unknown. Chemical probe profiling will be carried out *in vitro* and *in situ* to test functional predictions and potentially identify new proteins involved in exchange of this commodity. We believe this new chemical approach will reveal important vitamin-protein interactions in these systems, and importantly, be a technology universally applicable to more complex biological systems.



This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

172

Proteomic Characterization of Microbes and Microbial Communities

Mary S. Lipton[1]* (mary.lipton@pnnl.gov), Laurey Steinke,[2] Roslyn Brown,[1] Ronald J. Moore,[1] Therese R.W. Clauss,[1] Gordon Slysz,[1] Gordon A. Anderson,[1] David M. Ward,[3]* Christian Klatt,[3] Donald A. Bryant,[4] William P. Inskeep,[3] Mark Kozubal,[3] Zack Jay,[3] Jacob Beam,[5] and Ryan Jennings[3]

[1]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.; [2]Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha; [3]Montana State University, Bozeman; and [4]The Pennsylvania State University, University Park

Project Goals: The overarching goal of the PNNL Foundational Scientific Focus Area (FSFA) is to investigate fundamental scientific issues in microbial interactions—those that occur between different microbes as well as between microbes and their extracellular environment—to understand how microbial communities work collectively to carry out complex biogeochemical processes. This research component will develop and apply proteomic approaches to define these interactions, to track the flow of carbon from the autotrophic population to the heterotrophic population and to elucidate the metabolic control of energy flow and nutrient cycling in thermophilic microbial mats.

The challenges associated with the proteomic characterization of natural microbial mat communities growing in extreme environments are as varied as the types of environments. These challenges range from extremely small sample size and the presence of contaminating inorganic and organic materials in the mats to the difficulties in delineating which proteins and which organisms are responsible for the function that is present. To address these challenges, we are developing and applying methods for the extraction and analysis of proteins from the mats that include in situ stable isotope probing, advanced protein extraction methods, and subcellular fractionation methods in mixed populations.

Microbial autotroph- heterotroph associations are metabolically interactive, self-sustaining biological systems that are widespread in terrestrial and aquatic environments. Of specific interest to the FSFA program are phototrophic microbial mats found in alkaline siliceous hot springs in Yellowstone National Park (YNP), and chemoautotrophic iron and sulfur-oxidizing mats found in acidophilic, hyperthermophilic springs in YNP.

While the characterization of the microbial mats at Mushroom springs at YNP is limited by small samples sizes, the well annotated metagenome and relative ease of sample processing allow advanced analysis using *in situ* Protein Stable Isotope Probing (Pro-SIP) labeling strategies. Specifically, Pro-SIP has been applied to samples taken directly from the field with the immediate short-term exposure to ^{13}C-CO_2. Advanced analytics and informatics developed at PNNL have allowed for identification of taxa and associated proteins that are expressed and are actively taking up CO_2 during the incubation at the tested time point. A preliminary analysis of the first 10% of the data yielded 24 proteins validated by two or more unique labeled peptides. Since 14 were from *Roseiflexus RS-1*, 9 were from *Synechococcus* and only 1 was from *Candidatus Thermochlorobacter* aerophilum, it was determined that *Roseiflexus RS-1* is active in reducing CO_2 in the morning period of sunrise through the beginning of high light, and *Synechococcus* is active to a lesser extent in this time frame. During this time frame, these organisms are mainly synthesizing chaperones and transcription associated proteins.

In contrast to the phototrophic mats, the chemoautotrophic mats from YNP present different challenges for proteomic characterization. The relative high level of archaeal components and the high concentration of iron and sulfur in the mats confound cell lysis and protein extraction. To overcome these challenges, we have developed novel protein extraction techniques and have successfully detected a suite of iron, sulfur and arsenic oxidation proteins from the organisms present in the these communities. The meta-genome analysis suggests that the organism *Thermoproteales* (strain WP30) is present and further that this heterotrophic population respires on elemental sulfur and/or arsenate during growth on complex carbon sources. This genomic hypothesis is supported by the identification of the Mo-pterin proteins responsible for reduction of elemental sulfur and arsenate in these communities.

We will present the new and developing technologies for protein extraction, protein labeling and subcellular fractionation for organisms inhabiting natural microbial mats in extreme environments along with information on the

interactions among these organisms and their relationship to that environment.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

173

Integrating Geochemical, Microscopic, and –Omics Analyses to Understand Microbial Interactions in High-Temperature Chemotrophic Communities

William P. Inskeep[1]* (binskeep@montana.edu), Mark Kozubal,[1] Zack Jay,[1] Jacob Beam,[1] Ryan Jennings,[1] Hans Bernstein,[1] Laura Whitmore,[1] Ross Carlson,[1] Roslyn Brown,[2] Mary Lipton,[2] Helen Kreuzer,[2] James Moran,[2] Alice Dohnalkova,[2] Margie Romine,[2] James Fredrickson,[2] Susannah Tringe,[3] and Tanja Woyke[3]

[1]Montana State University, Bozeman; [2]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Wash.; and [3]DOE Joint Genome Institute, Walnut Creek, Calif.

Project Goals: To understand metabolic control of energy flow and nutrient cycling, and interactions among microbial community members in thermophilic chemotrophic microbial mats.

The advent of molecular tools and -omics technologies has provided opportunities for assessing the predominant and relevant indigenous organisms present in natural microbial communities, as well as their function within a connected network of different populations. High-temperature microbial communities are often considerably less diverse than mesophilic environments and are constrained by geochemical attributes such as pH, dissolved oxygen, Fe, sulfide, and or trace elements including arsenic and mercury. Consequently, one of the project goals is to understand microbial interactions among chemoautotrophic and heterotrophic members of high-temperature acidic Fe-oxidizing communities and sub-oxic elemental sulfur systems in Yellowstone National Park (YNP). Specific objectives of this work are to identify the predominant transcripts, proteins, metabolites and isotopic signatures associated with high-temperature microbial communities and to establish metabolic network models for different habitat types using consensus sequence assemblies of major phylotypes present in Fe(III)-oxide mats and sulfur sediments. Microbial community metabolic models derived from metagenome assemblies of individual populations are being used to test hypotheses regarding the role of specific phylotypes within each habitat type as well as microbial interactions occurring *in situ*.

Prior metagenome sequencing of high-temperature Fe-oxide and elemental sulfur systems from Yellowstone National Park (YNP) reveal communities dominated by thermophilic archaea and/or members of the deeply-rooted bacterial order Aquificales. Phylogenetic and functional analysis of metagenome sequence has provided an excellent foundation for hypothesizing the role of individual populations in a network of interacting community members, and for testing specific hypotheses regarding the importance of biochemical pathways responsible for material and energy cycling. For example, the predominant microbial populations present in acidic Fe-oxide microbial mats of Norris Geyser Basin (YNP) include chemoautotrophs such as *Metallosphaera yellowstonensis*, a member of the crenarchaeal order Sulfolobales. Pure-culture laboratory experiments have recently confirmed that this chemoautotroph can fix inorganic carbon (i.e. CO_2) using the 3-hydroxypropionate/4-hydroxybutyrate pathway while obtaining energy from the oxidation of ferrous Fe. Gene expression studies under different treatment conditions showed that *M. yellowstonensis* utilizes a novel terminal oxidase complex to oxidize Fe(II) (*fox* gene complex) while fixing inorganic carbon. Consequently, *M. yellowstonensis* is an important primary producer in high-temperature acidic Fe-oxide mats, and may provide a source of organic carbon for other heterotrophs present in the community.

Metagenome sequence assemblies show that the Fe-oxide mats also contain potential heterotrophs including Desulfurococcales and Thermoproteales-like populations. Functional analyses of genes belonging to these phylotypes suggest that these organisms degrade complex carbon sources, and that specific proteins may serve as a primary carbon and energy source. Recent isolation and characterization of a representative Thermoproteales organism (strain WP30) from YNP shows that this heterotrophic population respires on elemental sulfur and/or arsenate during growth on complex carbon sources. The Mo-pterin proteins responsible for reduction of elemental sulfur and arsenate have been identified and gene expression studies are underway to confirm the role of these novel proteins in community function. A deeply-rooted archaeal population has also been identified as a major community member in high-temperature Fe mats (referred to here as 'Novel Archaea Group 1'), and *de novo* sequence assemblies suggest that this organism is heterotrophic, potentially utilizing complex carbon sources produced by *M. yellowstonensis*. Iron depositional studies have been conducted to correlate Fe-oxidation rates with O_2 flux rates measured at the Fe-mat interface (using O_2-microelectrodes) as well as 16S rRNA gene copy-numbers of *M. yellowstonensis* (measured using quantitative PCR). Results show an excellent correlation between *M. yellowstonensis* copy-number and Fe-oxide deposition rate, especially during early stages of Fe-oxide mat development.

Early results from proteomic analyses of Fe-oxide mat samples confirm the importance of *M. yellowstonensis* and Novel Archaeal Group 1 populations in high-temperature Fe-oxidizing communities. Proteins identified in Fe-mat communities are being used to refine and improve metabolic models constructed using genome sequence. Pathway

specific processes are also being elucidated using isotope measurements focused on ^{13}C and ^{34}S of different chemical fractions. Additional proteomic and future transcriptomic results will be used to assess the importance of specific pathways, and will provide detailed information necessary to test hypotheses regarding the major microbial interactions that define community structure and function. Integration of genomic, proteomic, and metabolic information to understand autotroph-heterotroph interactions is tractable within high-temperature geothermal systems, in part due to the relative simplicity of the communities and the stability of key geochemical variables including pH, Fe, O_2 and dissolved sulfide/elemental sulfur.

This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

174

In Situ Correlated Molecular Imaging of Chemically Communicating Microbial Communities

R. Masyuko,[1] C.M. Driscoll,[1] **P.W. Bohn**[1]* (pbohn@nd.edu), J. Shrout,[1] and J.V. Sweedler[2]

[1]University of Notre Dame, Notre Dame, Ind. and [2]University of Illinois, Urbana-Champaign

Project Goals: (see below)

This project is exploring the potential of heterocorrelated mass spectrometric (MS) and confocal Raman microscopy (CRM) chemical imaging, as targeted to the problem of microbial/environmental processes. Specifically we are: (1) enhancing the functionality and performance of secondary ion mass spectrometry (SIMS) and laser desorption ionization (LDI) via tandem MS and improving the spatial resolution and analyte range by introducing a high-flux C_{60} source along with improved stage control and automated collection routines; (2) developing nanoparticle-enhanced correlated imaging with heightened spatial/temporal resolution and increased sensitivity; (3) developing a laboratory testbed for the *in situ* (natural) system to enable the simultaneous investigation of all components (bacteria-root-fungus) of a three-component rhizosphere model. Initially, we are examining the bacterium *Pseudomonas aeruginosa*, starting with the "relatively" simple case of *P. aeruginosa* group motility on idealized surfaces, imaging of homoserine lactones and surface remodeling with rhamnolipid and then transitioning to a more complex three-component system, composed of a bacterium (*P. aeruginosa*), a fungus (*Laccaria bicolor*) and model root derived from *Populus deltoides*. This poster will report on initial results based on chemometric approaches to Raman spectral devolution in complex matrices and enhancements to sensitivity in both LDI-MS and SIMS imaging.

175

From Genomes to Metabolomes: Studying Mechanisms of Interspecies Interaction Using the Archaeal System *Ignicoccus-Nanoarchaeum*

Richard Giannone,[1] Brian Bothner,[2] Valérie Copié,[2] Robert Hettich,[1] and **Mircea Podar**[1,3]* (mpodar@utk.edu)

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Montana State University, Bozeman; and [3]University of Tennessee, Knoxville

Few, if any, microbes live in functional or spatial isolation. The nature of the various types of inter-species interactions can be complex, ranging from competition to syntrophy and mutualism. Such relationships can impact keystone species and play a major role in energy and element cycles at scales that extend past ecosystem boundaries. There is limited understanding of fundamental mechanisms of interspecies recognition and communication, how they impact genome evolution, what genetic regulatory mechanisms control metabolic/energetic coupling between species in response to environmental factors. To address such questions we are using the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans*. With a combined genomic complement of less than 2000 genes and an obligate chemolithoautotrophic metabolism, this system represents one of the simplest specific microbial association and allows fundamental system level investigations and modeling of symbiosis. The integrated approach and the results of this research will be applied to investigations of more complex environmental systems.

A laboratory cultivation system for *Ignicoccus-Nanoarchaeum* has been established. The genomes of the two organisms are available and in addition we have sequenced two additional *Ignicoccus* species that do not interact with *N. equitans*. We have also performed an in depth proteomic analysis of *Ignicoccus-Nanoarchaeum* (1). Differences in the relative abundance of >75% of predicted protein-coding genes from both Archaea were measured to identify the specific response of *I. hospitalis* to the presence of *N. equitans* on its surface. A purified *N. equitans* sample was also analyzed for evidence of interspecies protein transfer. The depth of cellular proteome coverage achieved is amongst the highest reported for any organism. Based on changes in the proteome, *I. hospitalis* reacts to *N. equitans* by curtailing genetic information processing (replication, transcription) in lieu of intensifying its energetic, protein processing and cellular membrane functions. Using the information from initial studies we are now combining cultivation of these archaea under various settings with parallel transcriptomic, proteomic and metabolomic analyses. We are testing the specific hypothesis that the physical interaction between *Ignicoccus hospitalis* and *Nanoarchaeum* in laboratory cultures is induced and controlled by specific temporal gene expression, metabolic events and surface protein-protein interactions. We will identify candidate genes, proteins, and small molecules regulating the metabolic/energetic coupling

network shared by the two organisms, which will allow us to establish a model of symbiosis at the genomic and metabolic level. Using comparative and functional genomics in multi-species cultures with archaea that do not serve as hosts for *Nanoarchaeum*, we will independently test the mechanism of interaction between *I. hospitalis* and *Nanoarchaeum* and will derive an evolutionary genomics model for the development of a microbial symbiotic relationship.

Reference

1. Giannone RJ, Huber H, Karpinets T, Heimerl T, Küper U, Rachel R, Keller M, Hettich RL, Podar M. (2011). Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans-Ignicoccus hospitalis* relationship. PLoS One. 2011;6(8):e22942.

This research is funded by the U.S. Department of Energy, Office of Biological and Environmental Research under the Genomic Science Program (DOE-DE-SC0006654).

176

System-Scale Modeling of Mycorrhizal Symbiosis

Peter Larsen*[1] (plarsen@anl.gov), Geetika Trivedi,[2] Avinash Sreedasyam,[2] Leland Cseke,[2] and **Frank Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill. and [2]Department of Biological Sciences, University of Alabama, Huntsville

http://www.bio.anl.gov/molecular_and_systems_biology/proteins.html

Project Goals: The Argonne "Environment Sensing and Response" Scientific Focus Area (ESR-SFA) program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The overall objective of the ESR-SFA program is to identify the molecular basis of cellular transport and sensory pathways that mediate the response to environmental nutrients. This project furthers the objectives of this ESR-SFA by developing approaches that evaluate plant/fungal symbiosis by inferring molecular mechanisms from transcriptomic analysis of a laboratory *Laccaria bicolor*:aspen mycorrhizal system. Our approach generates testable biological hypotheses for the specific molecular mechanisms that drive nutrient exchange in ectomycorrhizal symbiosis. Increased understanding of how partners in plant/fungal symbiosis sense and respond to environmental parameters will identify the specific molecular metabolic and environmental response pathways for eukaryotic organisms in soil ecosystems.

The roots of many plant species have the ability to form symbiotic relationships with the community of soil fungi. In these relationships, mycorrhizal fungi provide mobilized nutrients from the soil, and plants provide photosynthetically-derived sugars. The plant:fungal symbiosis can provide growth and survival benefits, but the many details of the molecular exchanges are not completely understood. Deep RNA sequence analysis has revolutionized our access to the molecular activity of these communities, providing a system-scale perspective of symbiotic interactions at unprecedented resolution. However, no single method of transcriptomic analysis alone is sufficient to enable the inference of metabolite-space from transcriptomic-space. Identification of significantly expressed genes and mapping those genes onto known metabolic pathways generates a scaffold on which analysis of differential expression and predicted metabolic turnover can be accomplished. The results of this analysis are predictions of specific metabolic compounds and expressed protein activities that can be used to generate hypothesis-driven molecular biological experiments.

We generated a model of mycorrhizal metabolome that confirms prior biological knowledge and provides insight into the nutrient exchange process associated with mycorrhizal interaction. When *L. bicolor*:aspen mycorrhizae are cultured on media with inorganic nitrogen sources, *L. bicolor* is predicted to take up inorganic nitrogen from the media and synthesize complex nitrogenous compounds for its symbiotic plant partner. These predictions indicate *L. bicolor* is an active metabolic partner in mycorrhizal interactions in addition to passively extending the absorptive surface of aspen roots. Proposed transcriptomic experiments, examining mycorrhizal systems in the context of different nutrient environments, will uncover additional mechanisms of mycorrhizal metabolic interactions. Application of further transcriptomic analysis methods such as identifying gene splice variants expressed during different stages of mycorrhizal interaction, have the ability to provide even greater richness to the transcriptomic-generated models of the mycorrhizal metabalome. The approach applied to this ecto-mycorrhizal system can be generalized to other interacting systems in the rhizosphere.

177

Plant-Microbe Interfaces: Mining Genomic Signatures of Species-Specific Microbe Interactions in *Populus* Using High Density SNP Arrays and Whole Genome Resequencing

Wellington Muchero* (mucherow@ornl.gov), Jessy L. Labbé, Ranjan Priya, Lee Gunter, **Gerald A. Tuskan**, and **Mitchel J. Doktycz**

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Highly specialized plant-microbe interactions such as mycorrhizal symbiosis exhibit remarkable species specificity in economically important plants such as *Populus*. As such, molecular genomic signatures of these specialized interactions are expected to closely mirror the molecular genomic signatures of species differentiation, offering a manageable entry point in efforts to identify and characterize genetic elements essential for establishment of such interactions. In this study, we characterized genome divergence on Chromosome 1 of the *Populus* genome known to harbor major genetic determinants of *Populus–Laccaria* symbiosis as a test case for identifying candidate genes based on species-level genome divergence. We used a high-density single nucleotide polymorphism (SNP) array with a genome-wide coverage based on 34,130 probes to genotype three *Populus* species. Out of 3,716 SNPs on chromosome 1, we identified 229 SNPs that were not transferable between *P. trichocarpa* and *P. deltoides* or *P. fremontii*. These non-transferrable SNPs were over-represented in three regions on chromosome 1 representing putative zones of genome divergence during *Populus* speciation. In these regions, 25 genes with putative functions in plant-microbe interaction were identified. These included a phosphate transporter-related gene previously implicated in symbiotic phosphate transportation in arbuscular mycorrhizae, as well as a Vapyrin gene that was reported to be an essential factor in intracellular progression of arbuscular mycorrhizal symbiosis. Resequencing data for these genes will be used to characterize differences in gene structure between *P. trichocarpa* and *P. deltoides* that could potentially explain species-related *Populus–Laccaria* interactions.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

178

Plant-Microbe Interfaces: Emerging Research Directions

Jay Chen* (chenj@ornl.gov), Jessy Labbé, Jennifer Morrell-Falvey, Sarah Melton, Jim Guo, Udaya C. Kalluri, Dave Weston, Dale Pelletier, **Gerald A. Tuskan**, and **Mitchel J. Doktycz**

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The molecular events that lead to recognition and colonization of a host plant by beneficial bacteria or fungi are poorly understood. Ongoing research in this project is aimed at identifying molecular, genetic and cellular events in *Populus* involved in recognition and establishment of beneficial microbial interactions, and at identifying and isolating microbes associated with natural *Populus* ecosystems. Results from these ongoing studies have led to initiation of new research directions to enhance our understanding of plant-microbe interactions. New efforts are focused on elucidating the molecular mechanisms of microbial biofilm formation on plant roots, the role of helper bacteria in promoting beneficial plant-fungal interactions, and on development of a *Populus* protoplast transient expression system for the molecular and biochemical characterization of regulatory proteins.

The mutualistic association between plant roots and microbes form an environment that is ideal for biofilm formation, including sufficient moisture and nutrients, which are supplied by the plant host. This project is focused on

identifying the specific conditions in the plant rhizosphere that trigger biofilm formation in selected microbes. To this end, we have tested the ability of selected microbes to form biofilms under various growth conditions using a well-described biofilm formation assay. Data from multiple *Pseudomonas* strains indicate differential responses to additions in the growth media. For example, sucrose, a common constituent of plant root exudates, appears to induce biofilm formation in GM49. Likewise, the presence of high levels of phosphate induces biofilm formation in GM60 and GM67. These and other phenotypes will form the basis for selecting microbes on which to focus in-depth studies to elucidate molecular pathways involved in root colonization.

The soil is probably one of the most complex ecosystems in which plant–fungal–bacterial interactions operate. Mycorrhizal fungi are surrounded by complex microbial communities, which modulate the mycorrhizal symbiosis that impacts biomass production, defense against pathogens, and tree nutrition. Among this diverse microbial community, the so-called mycorrhiza helper bacteria (MHB) are thought to assist mycorrhiza formation and symbiosis. Since very little is known about the role of MHB in *Populus*-fungi interactions, this project is aimed at dissecting the signaling mechanisms underlying *Populus*–fungal–bacterial interactions. To this end, we have analyzed the morphological changes induced in *Laccaria bicolor* by several bacterial strains isolated from the *Populus* rhizosphere, and performed *Populus–L. bicolor*–bacteria co-cultures under controlled greenhouse conditions. We demonstrate that some bacterial strains influence *Populus–L. bicolor* colonization and have a clear beneficial effect. This study provides new insights into the mechanism of interaction between *Populus* and its complex microbial communities.

Transcriptomcis, proteomics and metabolomics have uncovered a suite of genes involved in *Populus*-microbe interactions. However, it remains unclear how these components interact with each other and are coordinated in the same signaling cascades to regulate some specific processes of plant-microbe interactions. Molecular and biochemical characterization of these genes is essential for elucidating the molecular mechanism underlying plant-microbe interactions. We have established a protocol for efficient isolation of protoplasts from *Populus* leaf mesophyll cells. Subsequently, we have established a protoplast transient expression system using *Populus* protoplasts. We demonstrated that *Populus* protoplasts respond to biotic and abiotic stimuli in a similar manner as that in intact plants. Furthermore, by using a series of fluorescent fusion proteins, we have established subcellular localization of various organelle markers. The *Populus* protoplast transient expression system will be used to study protein-protein interactions, protein-DNA interactions, post-translational modifications, protein degradation, and artificial microRNA-mediated gene silencing. Furthermore, together with gene reporters, this system will be used to examine the specific role of candidate genes in regulating plant-microbe interactions. The *Populus* protoplast transient expression system represents a much-needed system for research in the post-omics era.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

179

Plant-Microbe Interfaces: Revealing Regularities in Large Datasets with Biological Annotations Using Networks and Rules of their Association

Tatiana Karpinets[1]* (karpinetstv@ornl.gov), Byung Park,[2] Guruprasad Kora,[2] Mustafa Syed,[1] Christopher Schadt,[1] Gerald Tuskan,[1] Dale Pelletier,[1] **Edward Uberbacher**,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The expanse of experimental data produced by the Plant-Microbe Interfaces (PMI) project requires the development and application of novel computational tools for their processing. The data, collected from thousands of samples from diverse geographical locations in multiple PMI experimental campaigns, include a complex set of biological annotations, like phenotypic and genotypic characteristics of *Populus*, structure of fungal and bacterial communities in the tree rhizosphere and endosphere, and inherently heterogeneous soil properties. Combined in one table, this mash up of qualitative and quantitative data is challenging to analyze and interpret. To address this challenge we developed a computational framework that expands current approaches to viewing, visualizing, searching, and analyzing information in large databases with biological annotations. The approach is based on two novel concepts, the type-value format and the association network (Anet) that supplement the idea of association rules (Arules) produced by 'Apriori'. Type-value format simplifies computational processing, filtering and grouping of biological annotations by preserving their two-level structural organization in the transaction

records and further in Arules and Anets. For annotation types presented by quantitative data, such as genome size or GC content, quantities are replaced with their quality levels based on distributions of the quantities in the dataset records. Association network provides a way to link a diverse set of annotations directly, by the number of transactions where each pair of annotations co-occurred, and indirectly, by considering similarity between profiles of their co-occurrences with other annotations. Monte Carlo simulation is used to assess the significance of similarity by p-value. The resulting Anet is further analyzed and visualized at different levels of resolution, or p-value thresholds, using clustering and visualization techniques. In combination, Anets and Arules provide researchers a powerful tool to create a bird's-eye view of the collected information, to extract hidden biological regularities and to generate hypotheses for further experimental validation.

To test this framework we applied it to the analysis of metadata of sequenced prokaryotic genomes from the GenomeOnLine Database (GOLD). The overlapping structure of the data provides a good case study for the proposed framework. The generated Anet revealed a hidden structure in metadata of the prokaryotic organisms with three major clusters representing metadata of pathogens, environmental isolates and plant symbionts. The annotations clustered in each group represented a distinct signature profile of metadata for each group and showed a strong link between phenotypic, genomic and environmental features of the organisms.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

180

Plant-Microbe Interfaces: Dynamic Data Analytics Through Integrated Knowledgebase and LIMS

Guruprasad Kora[2]* (koragh@ornl.gov), Susan Holladay,[1] Tatiana Karpinets,[1] Mustafa Syed,[1] Christopher Schadt,[1] **Edward Uberbacher**,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn. and [2]Computer Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Data management, efficient tracking and effective visualization are essential aspects of any field-study related biological process. Each of these aspects pose numerous challenges as size and complexity of the data and field-study campaigns grow. Establishing a disciplined data management infrastructure at all levels of field-study campaigns enables effective life-cycle (data collection to analysis) management of any biological processes.

The Plant-Microbe Interfaces (PMI) project has developed an easy to use, web-based data management and tracking portal as part of their Knowledgebase for PMI field-studies that enables users to track and manage the entire life-cycle of the collected data. PMI LIMS and Knowledgebases are coupled together to form an integrated data analytics platform to perform statistical as well as data mining operations dynamically. The Knowledgebase augments data analytics infrastructure with scientific visualization techniques (visual data analysis). It enables users to use graphical representation of data as a means of gaining understanding and insight into the data. It helps researchers to comprehend spatial and temporal relationships between collected data. It provides efficient interactive techniques for researchers to focuses on exploratory, comparative analytics and visualization.

The PMI Knowledgebase has an established data acquisition workflow that enabled field-studies carried out in 2009-2011 to be effectively managed. Currently, the system supports various types of data from raw soil readings to tree specification to post-processed data like 454 sequences, isolates etc. The system supports various datatypes from image files, ascii-based text files, binary blobs etc. The workflow interface enables users to access the underlying data management layer with an easy to use and intuitive web interface. The interface seamlessly connects to laboratory-wide LIMS environment and makes day-to-day tasks like raw-data fetching and data summarization extremely efficient and easy. Furthermore, the portal provides a dynamic data analytics environment that facilitates users to perform standard statistical analysis on the LIMS stored data.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

181

Plant-Microbe Interfaces: Genotype-by-Environment Interactions Drive Root-Associated Microbiome Composition in Natural Populations of *Populus deltoides*

Christopher W. Schadt[1,2]* (schadtcw@ornl.gov), Migun Shakya,[1,2] Neil Gottel,[1] Hector Castro,[1] Zamin Yang,[1] Marilyn Kerley,[1] Gregory Bonito,[3] Dale Pelletier,[1,2] Tatiana Karpinets,[1] Guruprasad Kora,[1] Jessy Labbé,[1] Edward Uberbacher,[1,2] Wellington Muchero,[1] Gerald Tuskan,[1] Mircea Podar,[1,2] Rytas Vilgalys,[3] and **Mitchel J. Doktycz**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Genome Science and Technology Program, University of Tennessee, Knoxville; and [3]Department of Biology, Duke University, Durham, N.C.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Members of the genus *Populus* represent genetically diverse, ecologically widespread riparian species that are potential cellulosic feedstocks for biofuels, and the first woody plant species to have a genome sequence. The trees are also host to a wide variety of symbiotic microbial associations within their roots and rhizosphere. Thus they serve as an ideal model to study interactions between plants and microorganisms. However, most of our knowledge of microbial associations to date comes from greenhouse and young plantation-based trees; there have been few published efforts to comprehensively describe microbial communities of mature natural populations of *Populus*. We have compared root endophyte and rhizosphere samples collected from two dozen sites within watersheds/populations of *Populus deltoides* in Tennessee and North Carolina over multiple seasons. 454 pyrosequencing has been applied to survey and quantify the microbial community associated with *P. deltoides*, using primers targeting the V4 and V7-8 regions of the bacterial 16S rRNA gene and the D1 region of the fungal 28S rRNA gene. Genetic relatedness among the *Populus* trees was evaluated using 20 SSR markers chosen for distribution across all 19 linkage groups of the *Populus* genetic map. Soil physical, chemical and nutrient status, as well as tree growth and age characteristics were also evaluated. Root endosphere and rhizosphere communities have been found to be composed of distinct assemblages of bacteria and fungi with largely non-overlapping OTU distributions. Within these distinct endophyte and rhizosphere habitats, community structure is also influenced by soil characteristics, watershed origin and/or plant genotype, while observed seasonal influences have been minimal. We have also isolated over a thousand bacteria and fungi from these environments representing dominant community members *in situ.* Many of these isolates show distinct growth phenotypes with *Populus.* These findings indicate that the characteristics of the *Populus* root/soil environment may represent a relatively strong selective force in shaping endophyte and rhizosphere microbial communities and their functions.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

182

Plant-Microbe Interfaces: Characterization of IAA7.1 in *Populus* and its Role in Plant-Microbe Interactions

Poornima Sukumar[1]* (sukumarp@ornl.gov), Tatiana Karpinets,[1] Karuna Chourey,[2] Gregory Hurst,[2] Jay Chen,[1] **Gerald A. Tuskan**,[1] **Udaya C. Kalluri**,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division and [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Auxins are a class of plant hormones regulating many physiological and developmental processes. Auxin signaling involve a group of transcription repressor proteins known as AUX/IAA (Auxin/IAA induced), and another family of transcription factors that induce auxin induced gene

expression called ARFs (Auxin Response Factor). Both AUX/IAAs and ARFs have conserved protein family domains. Many studies in *Arabidopsis* have explored the individual roles of AUX/IAAs and ARFs. Our comparative bioinformatics analysis of the gene families in *Populus* and *Arabidopsis* shows that *Populus* typically has multiple orthologs (co-orthologs) for a given *Arabidopsis* gene. It is intriguing to understand whether the co-orthologs differ functionally in a more complex plant system such as *Populus*, with respect to the auxin signaling pathway and/or in establishing beneficial plant-microbe relationships. The present study examines the effect of knocking down two *Populus* co-orthologs (PtrIAA7.1 and PtrIAA7.2) of the *Arabidopsis* gene, IAA7- PtrIAA7.1 has a unique protein family domain that has thus far not been reported in any other member of the gene family, in any species. The domain is a tandem duplication of domain II, which is known to cause protein instability in other plant species. Physiological comparisons show that RNAi lines of PtrIAA7.1 have more severe phenotypes than those of PtrIAA7.2. PtrIAA7.1 mutants display reduced plant height, increased lateral branching, and reduced of apical dominance. Additionally, co-culture with *Piriformospora indica*, *Laccaria bicolor* strain S238N, and *Pseudomonas spp* strain GM30, resulted in decreased responses in the PtrIAA7.1 mutant compared to control plants, including reduced sensitivity to microbe induced changes in shoot and root growth and fresh weights. Protein profiling of shoot and root samples of PtrIAA7.1 revealed that shoots of PtrIAA7.1 lack a predicted importin protein belonging to a class of proteins involved in import of substrates into the nucleus. Additionally, PtrIAA7.1 mutant has lower abundance of ADP Ribosylation Factor proteins. Further experiments are underway to investigate the potential role of these differential expressed proteins in the auxin signaling pathway and their regulation during plant-microbe interactions.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

183

Plant-Microbe Interfaces: Sequencing of Twenty-One *Pseudomonas* Genomes and Twenty-Three Genomes from Diverse Bacteria Isolated from *Populus deltoides* Rhizospheres and Endospheres

Sagar M. Utturkar[1]* (sutturka@utk.edu), **Steven D. Brown**,[1,2] Dawn M. Klingeman,[2] Courtney M. Johnson,[2] Stanton L. Martin,[2] Miriam L. Land,[2] Tse-Yuan S. Lu,[2] Christopher W. Schadt,[1,2] **Dale A. Pelletier**,[1,2] and **Mitchel J. Doktycz**[1,2]

[1]Genome Science and Technology Program, University of Tennessee, Knoxville and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://pmi.ornl.gov
http://genome.ornl.gov/microbial/PMI/

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Plant-microbe interactions within the rhizosphere have been shown to have important roles in plant health and productivity (*1, 4*). As part of an ongoing effort to better understand the microbial communities associated with native *Populus deltoides* (Eastern cottonwood), we have under taken both cultivation independent and cultivation dependent assessments of microbial communities from *P. deltoides* rhizospheres and endospheres (*3*). Our goal is to understand the diversity of the *Populus* microbiome and to elucidate the metabolic and molecular mechanisms responsible for shaping the *Populus*-microbial interface.

We sampled *P. deltoides* at sites along the Caney Fork River in central Tennessee and at Yadkin River in North Carolina, USA. These sites represent ecotypes and soil conditions that are common to this region. Approximately 1,100 diverse bacterial strains were isolated from these sites over multiple seasons. The isolates comprise 7 classes and 85 genera of bacteria including, *Actinobacteria* (14%), *Bacilli* (17%), *Flavobacteria* (6%), *Sphingobacteria* (3%), and α- (22%) β- (16%) and γ- (22%) Proteobacteria. A number of our isolates were *Pseudomonas* species, which have considerable genetic and phenotypic variability and different members are pathogenic, biocontrol and plant growth promoting bacteria. Representatives were chosen for genome sequencing based on phenotypic traits and phylogeny. A total of 24 γ-Proteobacteria, 6- β-Proteobacteria, 10 α-Proteobacteria, 2-Bacteroidetes and 2-Bacilli were sequenced.

Paired-end DNA libraries with an average insert size of 500bp were created and draft genome data was generated using the Illumina (*2*) HiSeq2000 technology. CLC Genomics Workbench (version 4.7.1) and FASTQC (*6*) were applied to trim reads for quality sequence data. We analyzed the effect of quality based trimming and filtering of reads on genome assemblies and compared assembly outputs generated by the Velvet assembler (Version 1.1.04) (*7*) and the CLC Genomics Workbench software. In each case found the selection of high quality reads was a key step for successful assemblies. Quality trimming dramatically improved the assembly i.e. reduced the number of contigs

while maintaining the expected genome size. Velvet assemblies were further optimized by selecting different Kmer values and employing a scaffolding algorithm within the program. In the case of *Rhizobium* sp. PDO-076 a GS FLX shotgun dataset was also available, which was used to generate a hybrid assembly by combining 454 reads and shredded Velvet scaffolds using Newbler (Version 2.6). Initial assembly validation was performed based on parameters such as number of contigs, genome size, N50, contig length etc.

Draft genome sequences were annotated at Oak Ridge National Laboratory using the Microbial Genome Annotation Pipeline automated annotation pipeline at ORNL, which is based on the Prodigal gene prediction algorithm (*5*). Final validation was performed by annotating the assembled genome with ORNL genome annotation pipeline and comparing with known conserved sequences. We assembled diverse genomes that varied in size from 4.4MB to 10.7MB and had G+C contents in the range of 33% to 69%. The 4.9 Mb of *Phyllobacterium* genome was assembled in as few as 39 contigs with largest contig size of 1.4 Mb, while maximum number of contigs generated was the 778 for the 5.4 Mb *Caulobacter* genome. The N50 statistics varies from 14,172 to 506,356 bp with average value of 154,391 bp. These new genome sequences will allow more comprehensive comparisons for bacteria involved in plant-microbe interactions.

References

1. Bais, H. P., T. L. Weir, L. G. Perry, S. Gilroy, and J. M. Vivanco. 2006. The role of root exudates in rhizosphere interations with plants and other organisms. Annu. Rev. Plant Biol. 57:233-266.
2. Bennett, S. 2004. Solexa Ltd. Pharmacogenomics 5:433-438.
3. Gottel, N. R., et al. 2011. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. Appl. Environ. Microbiol. 77:5934-5944.
4. Hodge, A., G. Berta, C. Doussan, F. Merchan, and M. Crespi. 2009. Plant root growth, architecture and function. Plant Soil 321:153-187.
5. Hyatt, D., G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
6. Kircher, M., P. Heyn, and J. Kelso. 2011. Addressing challenges in the production and analysis of illumina sequencing data. BMC Genomics 12:382.
7. Zerbino, D. R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

184

Plant-Microbe Interfaces: Whole Genome Sequence Analysis of Bacterial Strains Isolated From the *Populus* Microbiome

Dale A. Pelletier[1,2]* (pelletierda@ornl.gov), Tse-Yuan Lu,[1] Christopher W. Schadt,[1,2] Tatiana Karpinets,[1] Sagar Utturkar,[1,2] Dawn Klingeman,[1] Steven D. Brown,[1,2] Patricia Lankford,[1] Marilyn Kerley,[1] Amy Schaefer,[3] Caroline S. Harwood,[3] and **Mitchel J. Doktycz**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Genome Science and Technology Program, University of Tennessee, Knoxville; and [3]Department of Microbiology, University of Washington, Seattle

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Populus trees are host to a variety of microorganisms within their endosphere and rhizosphere that can have positive effects on the host. Our goal is to understand the phylogenetic and functional diversity within the *Populus* microbiome and to elucidate the metabolic and molecular mechanisms responsible for shaping the *Populus*-microbial interface. To begin to untangle this complex ecosystem, we isolated bacterial strains native *Populus deltoides* roots collected in central Tennessee and North Carolina. A diverse array of bacterial strains (>1000) comprising some 7 class and 89 genera of bacteria were isolated from the rhizosphere (529) and root endosphere (558). The isolates comprise Actinobacteria (15%), Bacilli (17%), Flavobacteria (6%), Sphingobacteria (3%), and a- (22%) b- (15%) and g- (22%) proteobacteria. In order to explore potential metabolic and physiological diversity present within *Populus* microbiome isolates, we performed whole genome sequencing on a subset of 43 bacterial isolates. Isolates where chosen for sequencing based on a variety of factors including abundance in native ecosystems, ability to colonize plants, both microbial and plant phenotyping and physiological properties. Sequence data was generated by Illumina HiSeq2000 paired-end sequencing of 500 bp insert libraries. Sequence reads were assembled using Velvet and annotated by the ORNL automated

genome annotation pipeline. The initial metabolic analyses of the sequenced strains targeted a comparison of KEGG orthologous groups. Metabolic activities of the organisms were predicted using the same KEGG annotation pipeline and then analyzed using network tools to identify clusters of KEGG orthologous groups that are specific for each isolate or group of isolates. The analysis revealed metabolic diversity of the isolates. Metabolic reconstructions generated for isolates provided further insight into the nature of metabolic interactions of the isolates with the plant host.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

185

Plant-Microbe Interfaces: Dynamics of Bacterial Microbiome of *Populus deltoides*

Migun Shakya,[1,2]* (shakyam@ornl.gov), Neil R. Gottel,[3] Hector F. Castro,[1] Zamin Yang,[1] Marilyn Kerley,[1] **Mircea Podar**,[1,2] Mitchel J. Doktycz,[1,2] and **Christopher W. Schadt**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Genome Science and Technology Program, University of Tennessee, Knoxville; and [3]Department of Microbiology, University of Texas, Austin

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The root-rhizosphere interface of *Populus* is an ideal model to study the interaction between plants and microorganisms. In our recent paper (Gottel et al. 2011), we characterized microbial communities from root endophytic and rhizospheric habitats of *P. deltoides* in mature, natural trees. However the relative effects of drivers of community composition such as soil and environmental properties, seasonal variation, and host genotype that shape these communities remains unclear. To address these issues, we carried out 454 pyrosequencing using primers targeting V7-V9 region of 16S bacterial SSU rRNA from samples collected from the rhizosphere and endosphere of two-dozen sites distributed across watersheds in North Carolina and Tennessee over two seasons. From a total of 686,384 high-quality, denoised, non-chimeric sequences, we identified key bacterial taxa associated with the endosphere and rhizosphere. Proteobacteria consistently dominated both environments, while Actinobacteria were found across all samples and in some endosphere samples they replaced Proteobacteria as the dominant taxa. More than 24,000 unique bacterial OTUs were detected however 21,487 OTUs were exclusive to rhizosphere, while only 2,598 OTUs were found in endophyte samples. We did not observe a clear distinction in the community composition of samples collected over multiple seasons, but variations associated with location, soil properties, and plant genotypes were observed. Our data indicate that the endophyte associated bacterial community has low diversity, but is highly variable from tree to tree. While rhizosphere associated bacterial communities are highly diverse community makeup is very similar from tree to tree. This study is one of the more comprehensive analyses of bacterial communities in the endosphere and rhizosphere of *P. deltoides* to date and highlights the complexity of the bacterial diversity associated with mature trees in natural systems.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

186

Plant-Microbe Interfaces: Quorum Sensing Systems are Prevalent in the *Populus* Microbiome

Amy L. Schaefer*[1] (amyschae@uw.edu), **Dale A. Pelletier**,[2,3] Sagar Utturkar,[2,3] Steven Brown,[2,3] Colin R. Lappala,[1] Ryan Morlen,[1] Daniel B. Mueller,[1] E. Peter Greenberg,[1] and **Caroline S. Harwood**[1]

[1]Department of Microbiology, University of Washington, Seattle; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [3]Genome Science and Technology Program, University of Tennessee, Knoxville

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of

these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

As part of the ORNL Plant-Microbe Interfaces Science Focus area, we are characterizing the natural diversity of microbial associates of *Populus* and elucidating the molecular mechanisms by which these organisms interact. We sampled a population of *P. deltoides* as it occurs along the Caney Fork River in Tennessee in 2009. Analysis of 16s rRNA sequences indicates the *Populus* bacterial communities are dominated by Acidobacteria, Alphaproteobacteria, and Gammaproteobacteria and the Proteobacteria are the predominant group isolated from *Populus* endophyte and rhizosphere samples (Gottel et al, Appl Environ. 77:5934). Many Proteobacteria use acyl-homoserine lactone (AHL) signals for cell density-dependent gene regulation, in a process known as quorum sensing and response. LuxI-type proteins synthesize small, diffusible AHL signals that function with LuxR-type signal receptors to control gene expression. Most known AHLs possess a fatty acyl side chain, derived from fatty acid biosynthesis, of varying side chain length and substitution. Recently we discovered novel AHL-type signals (*p*-coumaroyl-HSL, cinnamoyl-HSL, and isovaleryl-HSL) made by several soil- and plant-associated Alphaproteobacteria. This suggests that there may be additional novel HSL-type signals made by bacteria.

We initially screened ~120 Proteobacteria isolated from *P. deltoides* for AHL production and found >80% Alphaproteobacteria and >20% of the Gammaproteobacteria isolates to be positive. This suggested AHL signaling is prevalent in *Populus* microbial communities. When we examined the entire genome sequence of 40 of the Proteobacterial isolates, AHL signaling genes were even more prevalent than we predicted from AHL screening results as at least one *luxI*- or *luxR*-type gene was present in 10/10, 3/6, and 24/24 of the sequenced Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria genomes, respectively. Interestingly, sequence analysis of some of the *luxI*-type genes suggests the potential for novel AHL-type signals. Sequence analysis of the *luxR*-type genes suggests some receptors proteins may be responsive to an unknown plant-derived compound, as has reently been reporter for bacteria associated with certain crop plants (Subramoni et al, Appl Environ. 77:4579).

AHL quorum sensing often controls the production of "public goods" such as antimicrobials and exoezymes, as well as aggregation factors and conjugal transfer processes. In order to define the AHL regulon of a particular bacterium, mutants in either the *luxI*- or *luxR*-type genes are often constructed and analyzed relative to wild-type. However, not all AHL-producing bacteria are genetically tractable. To examine AHL-regulons in bacteria without constructing AHL-mutants we have demonstrated that purified AiiA lactonase, an enzyme that hydrolyzes the HSL ring of AHL signals, can be added to bacterial cultures to inhibit AHL-regulated phenotypes and gene expression. Using this enzyme we are defining the AHL-regulons of *Populus*-associated bacteria, by using RNAseq (Hirakawa et al, J Bacteriol. 193:2598).

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

187

Plant-Microbe Interfaces: *Populus deltoides* Supports Distinct Fungal and Bacterial Root Associates From Other Ectomycorrhizal Tree Hosts

Gregory M. Bonito[1]* (gmb2@duke.edu), **Christopher W. Schadt**,[2] Gerald A. Tuskan,[2] Mitchel J. Doktycz,[2] and **Rytas Vilgalys**[1]

[1]Biology Department, Duke University, Durham N.C. and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Populus deltoides is a common riparian tree species in areas of southeastern North America and is largely dependent on flooding for seedling recruitment. This species is unique from co-occuring tree species in that it forms root associations with both arbuscular and ectomycorrhizal fungal species. *Populus deltoides* also harbors both bacterial and fungal endophyte communities within its roots. To address the influence of edaphic or genotypic factors on the structuring of rhizospheric assemblages, we carried out a series of trap-plant experiments in growth chamber environment using rooted cuttings of *P. deltoides* and two other tree species (*Quercus alba* and *Pinus taeda*) grown in soils from our field sites. We used 454 multiplex amplicon pyrosequencing to characterize rhizospheric fungal and bacterial diversity for each plant in the experimental treatments. Specific objectives of this research were to:

1. Determine the effect of host species (*P. deltoides*, *Q. alba*, *P. taeda*) on the structuring of rhizosphere bacterial and fungal communities

2. Determine the influence of *P. deltoides* genotype on the structuring of rhizosphere bacterial and fungal communities
3. Determine the effect of different field soil inoculum on bacterial and fungal rhizosphere communities of individual host species and *P. deltoides* genotypes

Field soils from our ORNL *P. deltoides* research sites in NC and TN were used as the source of microbial inoculum in these trap-plant studies. Cuttings from different *P. deltoides* genotypes were planted into a 50% mixture of sterile sand and field soil. For the host treatment we grew *P. deltoides*, a *P. deltoides x P. trichocarpa* hybrid, oak (*Quercus alba*) and pine (*Pinus taeda*) in the same soil. For the soil treatment a single *P. deltoides* genotype was grown in multiple soil types. Plants were harvested after five-months of growth. Soils were washed off the root systems, roots visually assessed for ectomycorrhizas and fungal infection, and samples of bulk roots used for DNA extraction. The fungal community from each plant was sequenced at the ITS and LSU rDNA regions using the fungal specific primers ITS1f/ITS4 and LROR/LR3. Arbuscular mycorrhizae were preferentially amplified selectively with the primer set AML1 and a modified AML2 primer. Bacterial 16S rDNA primers that amplify across the V4 region and discriminate against plastid DNA were used to compare rhizosphere bacterial communities in selected samples. Pine and oak seedlings showed high ectomycorrhizal colonization (>80%), while most of the *P. deltoides* genotypes had low ectomycorrhizal colonization (<30%). This observation was verified through sequence-based assessments, with more than 75% of the sequences from *Populus* belonging to endophytic fungi and over half of those from oak and pine belonging to ectomycorrhizal fungi. The most abundant ectomycorrhizal fungi on *P. deltoides* belonged to the genera *Peziza, Inocybe* and *Hebeloma*. The *Peziza* and *Inocybe* taxa also co-occurred on oaks and pines. Ectomycorrhizal species of *Tuber* and *Laccaria* were recovered from both *Populus* and oak roots. Although *Populus* hosted fewer ectomycorrhizal taxa than either oak or pine, a more species-rich assemblage of endophytic fungi was detected in *Populus* roots. The total richness of root associated fungi (238) and bacteria (283) taxa was significantly greater for *Populus* than oak (175-f;184-b) or pine (157-f;185-b). For the bacteria, *Populus* was characterized by a higher relative abundance of *Actinobacteriales* and *Sphingobacteriales* than were oak or pine, and a lower abundance of *Rhizobiales* and *Berkholderiales.* Nonetheless, a core set of fungal (43) and bacteria (103) taxa were shared between the three hosts. There were some minor responses of bacterial and fungal communities to *Populus* genotype. In particular, the *P. deltoides x P. trichocarpa* hybrid had a wider spread across species space (based on PCA ordination) and was represented by a greater frequency and relative abundance of ectomycorrhizal taxa (e.g. *Inocybe, Tomentella, Hebeloma*). Arbuscular mycorrhizal fungi belonging to the Glomerales (*Glomus* sp.) and Paraglomales (*Paraglomus* sp.) were present in all genotypes, but sequences belonging to the Diversisporales were only detected in the hybrid genotype. Soils differed significantly in their effect on microbial communities. For example, a novel species of *Atractiellales* was the most abundant species in some assayed soils, but was absent in other soils. Similar patterns were observed for other microbial taxa. While we did detect fungi and bacteria in our negative controls, these clustered tightly in ordination space and showed little overlap with taxa from the experimental treatments. In conclusion, the microbiota associated with roots of *P. deltoides* is diverse and unique from oak and *pinus*, and appears to be structured both by the microbial inoculum available in soils and (to a lesser extent) plant genotype.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

188

Plant-Microbe Interfaces: Genome-Wide Identification of *Populus* Small Proteins Responsive to Mycorrhizal Symbiosis

Ting Li,[1] Jonathan Plett,[2] Chu-Yu Ye,[1] Gregory B. Hurst,[3] Jessy L. Labbé,[1] Wellington Muchero,[1] Priya Ranjan,[1] Hengfu Yin,[1] Gerald A. Tuskan,[1] Timothy J. Tschaplinski,[1] **Francis Martin**,[2] and **Xiaohan Yang**[1]* (yangx@ornl.gov)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Ecogenomics of Interactions Laboratory, UMR 'Interactions Arbres/Micro-Organismes', INRA-Nancy, France; and [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The mycorrhizal symbiosis, representing the most widespread plant-microbe association, offers various benefits including 1) enhancing carbon sequestration in terrestrial ecosystems, 2) increasing nutrient availability, 3) remediating degraded soils and 4) improving water use efficiency. All of these beneficial aspects make plant-mycorrhizal association an excellent strategy for improving the sustainability of bioenergy feedstock production. However, our knowledge about the molecular mechanism underlying mycorrhizal

* Presenting author

symbiosis is still very limited. To address this limitation, we performed a genome-wide analysis of *Populus* genes in response to *Laccaria* inoculation (Fig. 1). We identified 1,282 transcripts differentially expressed during mycorrhizal development, among which 764 novel transcripts have not been documented in the current version of *Populus* genome annotation. About 11% of the 1,282 differentially-expressed transcripts encode small proteins of <100 amino acids in length. Gene ontology analysis revealed that mycorrhizal symbiosis between *Populus* and *Laccaria* involves different sets of genes over the time course of symbiosis development (Table 1). A large number of protein sequences encoded by these transcripts were predicted to be located in the nuclei, suggesting that they may play roles in gene expression regulation. Interestingly, many protein sequences were predicted to be located in both nuclei and extracellular space, implying that they are putative signaling proteins responsible for communication between *Populus* and *Laccaria.* Also, we found that 11 transcripts were up-regulated through the time course of mycorrhizal development, suggesting that they are important to the maintenance of plant-fungus interaction. Since *P. trichocarpa* better associates with *Laccaria* when compared to *P. deltoides*, we examined the differences in the DNA-seq coverage of differentially expressed genes (based on genome resequencing data) between the two *Populus* species. We identified 10 genes that are present in *P. trichocarpa* but absent in *P. deltoides*. Lastly, we performed homology search in 24 other sequenced plant genomes ranging from algae to moss to angiosperm and found that only about 40% of the *Laccaria*-responsive genes in *Populus* have homologs in other plant species, suggesting that lineage-specific molecular mechanism may play an important role in regulating plant response to mycorrhizal fungi.

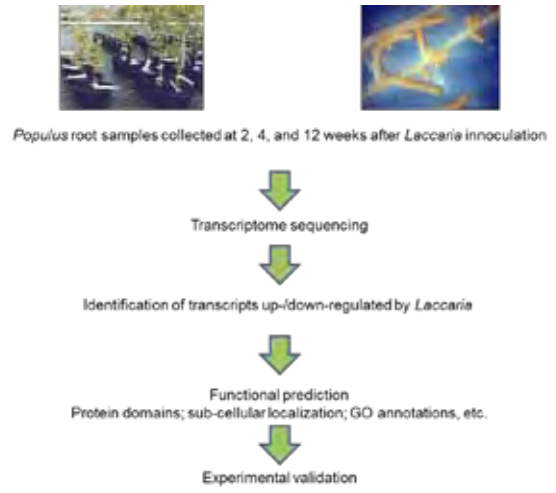


Figure 1. A pipeline for discovery of *Populus* small proteins mediating plant-fungus interaction.

Table 1: Biological processes enriched in up-/down-regulated *Populus* genes at 2, 4, and 12 weeks after *Laccaria* inoculation.

	2 weeks	**4 weeks**	**12 weeks**
Up regulation	Response to stimulus	Macromolecule metabolic process	Nitrogen compound metabolic process
Down regulation	Nitrogen compound metabolic process	Nitrogen compound metabolic process	Biosynthetic process

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

189

Plant-Microbe Interfaces: Extending Single Plant-Microbe Co-Expression and Metabolic Networks to Community Scales

David J. Weston (westondj@ornl.gov), Dale A. Pelletier, Abhijit A. Karve,* Jennifer L. Morrell-Falvey, Timothy J. Tschaplinski, Sara S. Jawdy, Lee E. Gunter, Nancy L. Engle, Jessy L. Labbé, Tse-Yuan S. Lu, Chris W. Schadt, **Gerald A. Tuskan**, and **Mitchel J. Doktycz**

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Colonization of plants by nonpathogenic *Pseudomonas fluorescens* strains can confer enhanced defense capacity against a broad spectrum of pathogens. Few studies, however, have linked defense pathway regulation to primary metabolism and physiology. In this study, physiological data, metabolites and transcript profiles are integrated to elucidate how molecular networks initiated at the root – microbe interface influence shoot metabolism and whole-plant performance. Experiments with *Arabidopsis thaliana* were performed using the newly identified *Pseudomonas fluorescens* GM30 or *P. fluorescens* Pf-5 strains. Co-expression networks indicated

that Pf-5 and GM30 induced a subnetwork specific to roots enriched for genes participating in RNA regulation, protein degradation and hormonal metabolism. In contrast, only GM30 induced a subnetwork enriched for calcium signaling, sugar and nutrient signaling and auxin metabolism, suggesting strain-dependence in network architecture. In addition, one subnetwork present in shoots was enriched for genes in secondary metabolism, photosynthetic light reactions and hormone metabolism. Metabolite analysis indicated that this network initiated changes in carbohydrate and amino acid metabolism. Consistent with this, we observed strain-specific responses in tryptophan and phenylalanine abundance. Both strains reduced host plant carbon gain as estimated by net photosynthesis, yet provided a clear fitness benefit when plants were challenged with the pathogen *Pseudomonas syringae* DC3000.

The trade-off between host carbon cost under optimal conditions and fitness benefit under pathogen attack brings to question how plant-microbe interactions are perceived and initiated within natural systems harboring complex microbial communities. Does the host, for example, favor carbon partitioning to microbial associates that in turn confer a fitness advantage? Alternatively, carbon acquisition from the host may simply be driven by the ability for the microbe to evade host plant defense. To begin to address these questions, we have developed a microcosm system providing axenic conditions in a realistic soil environment that is conducive to whole plant phenotype imaging and multi-omic sample collections. We are exploiting the use of whole-genome sequences from 43 microbial isolates collected from the *Populus* microbiome (see Pelletier poster) to create synthetic communities. Within this system, we are currently testing whether host plant genotype and nutritional state feedback to influence microbial community structure. In addition, we are testing whether the selected microbial community confers a beneficial host phenotype through community decomposition studies.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

190

Plant-Microbe Interfaces: Proteomics at the Plant-Microbe Interface

Gregory B. Hurst[1]* (hurstgb@ornl.gov), Karuna Chourey,[1] Xiaohan Yang,[2] Ting Li,[2] Chongle Pan,[1] Zhou Li,[4] Rachel M. Adams,[4] Robert L. Hettich,[1] Patricia K. Lankford,[2] Keiji G. Asano,[1] Tatiana V. Karpinets,[2] Udaya C. Kalluri,[2] Poornima Sukumar,[2] David J. Weston,[2] Sara M. Allen,[2] Dale A. Pelletier,[2,4] Jonathan Plett,[3] Francis Martin,[3] Timothy J. Tschaplinski,[2] **Gerald A. Tuskan**,[2] and **Mitchel J. Doktycz**[2,4]

[1]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Ecogenomics of Interactions Laboratory, INRA-Nancy, France; and [4]Genome Science and Technology Program, Knoxville, Tenn.

http://PMI.ornl.gov

Project Goals: Understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as an initial experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The proteomes and metaproteomes of plants, bacteria, and fungi offer insights into the mechanisms that these organisms employ at plant-microbe interfaces. We have initiated proteomics studies of several organisms involved in these interfaces, with an emphasis on those involving *Populus* trees, to identify processes in which proteins are important for establishment and maintenance of the interface.

-Proteomics of *Laccaria bicolor* mycelia We have performed "shotgun" proteomics of mycelia from the ectomycorrhizal fungus *Laccaria bicolor* grown under a variety of conditions, including different fungal strains, media compositions, cold and heat stresses, and proximity to roots of *Populus* plants. The average number of *L. bicolor* proteins identified per sample was ~1200. Approximately 440 proteins were detected in at least one replicate of each condition, establishing a "core" proteome for *L. bicolor*.

-Analysis of small proteins encoded by small genes in bacterial and plant species To complement work pioneered by Xiaohan Yang to identify small-protein-encoding genes (see poster by Li et al.), we are developing proteomics methods to target small proteins from plant tissues. After evaluating several methods for enriching small proteins in an *E. coli* model system, we applied the most promising technique (in-gel digestion of low molecular weight regions from 1D SDS-PAGE separations) to studying small proteins in *Arabidopsis* roots and shoots. Of a total of >5000 proteins identified from unfractionated extracts and small protein fractions, a few tens of proteins were identified exclusively in gel fractions corresponding to molecular masses below 20 kilodaltons. Evidence for expression of these proteins supports improved annotation of the corresponding small genes, and provides candidates for further studies of biological function.

-Proteomics studies of roots from field-sampled mature Populus trees The ability to study the root proteome of plants is an important first step towards studying plant-microbe inter-

actions in the rhizosphere. Extraction of proteins from plants is technically challenging, and further complications arise when the samples are obtained from mature tree roots. We have identified an approach that reliably yields proteins from roots of naturally occurring *Populus* trees, sampled during PMI collection trips to the Yadkin River in North Carolina. We typically identify >1000 proteins from each root sample, despite heterogeneity in morphology, differences in location, soil type, etc.; from 4 individual trees, we have evidence for ~2600 proteins to date (with some redundancies due to gene duplications, etc.). A "core" proteome that is common to all these root samples contains several hundred proteins. Analyses of functional categories and subcellular locations are ongoing.

-Comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics Using a model soil bacterial species, *Pseudomonas putida* F1, we have compared figures of merit such as depth of proteome coverage, quantification accuracy, precision, and reproducibility for several quantitative proteomics methods using a high-performance hybrid mass spectrometer, the LTQ Orbitrap Velos. Each approach has particular merits, and the final choice of approach depends on the requirements of the experiment at hand. Our results indicate that isobaric chemical labeling has the highest quantification precision, label-free quantification provides the largest number of protein identifications, and metabolic labeling is intermediate in both measures.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

191

Nutrient Cycling for Biomass: Interactive Proteomic/Transcriptomic Networks for Global Carbon Management Processes within Poplar-Mycorrhizal Interactions

Leland J. Cseke[1]* (csekel@uah.edu), Frank M. Collart,[2] Peter E. Larsen,[2] Stephen Barnes,[3] and Helen Kim[3]

[1]University of Alabama, Huntsville; [2]Argonne National Laboratory, Lemont, Ill.; and [3]University of Alabama, Birmingham

Project Goals: This project will facilitate the development of system-scale models of the symbiotic interaction between ectomycorrhizal fungi (such as *Laccaria bicolor*) and tree species (such as poplar) in response to environmental nutrient availability/biochemistry.

The experimental plan will integrate multiple "omics" approaches to model ectomycorrhizal regulatory networks and metabolic pathways that are predictive of atmospheric carbon sequestration in the form of plant and/or subsurface fungal biomass. The project will test the hypothesis that essential regulatory and metabolic mechanisms can be inferred from transcriptomic and proteomic changes that occur at the mycorrhizal interface in response to environmental nutrient availability. Guided by abundant genome sequence and ongoing transcriptomic input, this hypothesis will be addressed using modern protein analytic approaches to fill the gap in our understanding of how mycorrhizal metabolic and regulatory processes at the transcript-level translate to nutrient uptake, carbon management and ultimate net primary productivity of plants in the environment. Specifically, we make use of targeted as well as discovery-based proteomics, biochemical assays and ChIP-seq characterization of carbon, nitrogen and phosphorous regulators to identify symbiosis-specific molecular mechanisms that control plant carbon management and allocation. These molecular mechanisms regulate nutrient cycling, accumulation of plant and fungal biomass, and ultimately are important to forest management and atmospheric carbon sequestration.

192

Global Gene Expression Profiling of Switchgrass Following Inoculation with *Burkholderia phytofirmans* Strain PsJN

Alejandra Lara-Chavez[1]* (alejandra.lara@ialr.org), Seonhwa Kim,[1] Scott Lowman,[1,2] Yuhong Tang,[4] Barry Flinn,[1,2,3] John Seiler,[3] Jerzy Nowak,[2] and **Chuansheng Mei**[1,2,3]

[1]Institute for Sustainable and Renewable Resources, Institute for Advanced Learning and Research, Danville, Va.; [2]Departments of Horticulture and [3]Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg; and [4]Genomics Core Facility, The Samuel Roberts Noble Foundation, Inc., Ardmore, Okla.

http://www.ialr.org/images/stories/research/isrr/meiprojectreport9-28-11.pdf

Project Goals: 1) Global gene expression profiling induced by switchgrass bacterization with the beneficial bacterial endophyte *Burkholderia phytofirmans* strain PsJN and 2) Identification of key genes from global gene expression profiling and a study of their functions.

Switchgrass is one of the promising bioenergy crop candidates for the U.S. It gives relatively high biomass yields and can grow on marginal lands. However, the biomass yield varies from year to year and from location to location. The overall goal of the project is to develop a low input and sustainable switchgrass feedstock production system utilizing beneficial bacteria endophytes. Previous results on the inoculation of switchgrass lowland cultivar Alamo with *Burkholderia phytofirmans* strain PsJN indicated a significant increase of growth under *in vitro*, growth chamber and greenhouse conditions. However, no beneficial responses were recorded with the upland cultivar Cave-in-Rock. In order to explore this genotype effect further, a comparative global gene expression profiling was conducted both

cultivars following PsJN inoculation using the DOE-funded switchgrass cDNA microarrays. Ten-day old seedlings were inoculated with PsJN (0.5 at OD_{600}) and tissue samples collected at 0 (prior to inoculation), 0.5, 2, 4 and 8 days after inoculation, with three biological replicates. Non-inoculated controls were treated with PBS buffer and sampled as above.

Based on the microarray data analyses, 50 candidate genes that exhibited significant differences in the expression levels between PsJN-inoculated Alamo and Cave-in-Rock were chosen for further study. These genes are being subjected to secondary verification using qPCR. So far, 20 out of the 50 have been verified. From these 20 genes, five key genes representing glutathione S-transferase, calmodulin-related calcium sensor protein, an EF-hand transcription factor, histidine-containing phosphotransfer protein and a zinc-finger protein have been chosen for further functional studies using overexpression and RNAi knockout/knockdown techniques. Overexpression constructs for glutathione S-transferase and calmodulin-related calcium sensor genes were also introduced into switchgrass embryogenic callus, and plants will be regenerated and tested for endophyte PsJN responses.

Further analysis of the microarray data focused on groups of genes that are up-regulated in Alamo and down-regulated in Cave-in-Rock at each sampling point. We identified approximately 1947, 877, 402 and 1140 genes that displayed this pattern at 0.5, 2, 4 and 8 days after inoculation, respectively. The majority of genes showing differences are annotated as "expressed proteins" and "unknown proteins". We are currently focused on transcription factor genes, such as AP2 domain, MYB family, F-box domain, and zinc finger protein. Further studies of these transcription factor gene functions are listed in Table 1.

This research is supported by the Office of Science (BER), U.S. Department of Energy.

Table1. Expression level changes of transcription factor genes of interest in Alamo and Cave-in-Rock at 0.5, 2, 4 and 8 days following inoculation with PsJN, compared with expression level at 0 day, respectively.

ID probe	Annotation	Alamo				Cave-in-Rock			
		0.5	2	4	8	0.5	2	4	8
AP13ITG55712_at	AP2 domain	1.71	1.48	2.14	2.80	0.05	0.05	0.07	0.07
AP13ITG63524RC_s_at		2.27	1.75	2.59	2.29	0.89	0.68	0.79	1.14
AP13CTG22494_at	bZIP	1.88	3.58	3.03	1.80	1.27	1.48	1.51	0.95
AP13ITG54829_at		2.62	2.05	2.79	1.71	1.45	1.40	1.39	1.68
AP13CTG24092_at	MYB family	1.52	1.68	2.06	1.24	0.95	0.94	0.91	0.98
KanlowCTG34263_at		1.24	2.09	5.46	4.58	0.71	0.93	1.42	2.12
KanlowCTG22073_s_at		2.25	0.94	0.57	0.52	1.36	1.37	1.26	1.24
AP13ITG65291_at	F-box domain	1.53	2.03	2.26	2.88	1.15	1.06	0.79	0.98
KanlowCTG42852_s_at		1.20	1.70	2.13	2.15	0.75	0.77	0.66	0.68
AP13ITG41289_at		1.18	1.60	2.07	1.83	0.32	0.27	0.29	0.33
AP13ITG57608_s_at	RING-H2 finger	1.09	2.28	2.49	2.81	0.77	0.95	0.88	0.96
AP13ITG69131RC_at	zinc finger, C3HC4 type	1.56	1.76	2.07	2.28	0.74	0.62	0.67	0.71
AlamoCTG04292_s_at		2.26	1.22	1.16	1.11	1.65	1.95	1.89	1.90
AP13CTG19863_at	TFs having WRKY and zinc finger domain	3.13	1.94	1.70	1.69	0.12	0.11	0.17	0.17
AP13CTG44559_s_at		1.68	2.53	4.58	4.20	0.03	0.04	0.05	0.06
AP13.12336.m00003_s_at	No apical meristem	3.60	1.55	0.83	0.90	3.80	4.31	4.85	3.94
KanlowCTG46205_s_at	Transcription elongation factor	3.71	2.02	1.88	1.17	0.77	0.70	1.29	1.67
AP13CTG09371_s_at	zinc finger	2.53	1.30	0.89	0.73	1.15	1.82	1.96	1.95
AP13ITG48832_s_at	AT hook motif	2.73	1.38	0.67	0.45	1.20	1.33	1.35	1.19

193

ENIGMA Biotechnology: Systems Approaches to Studying Microbial Communities

Richard Baran,[1] Benjamin Bowen,[1] Morgan Price,[1] Christopher Petzold,[1] Simon Allen,[3] Susan J. Fisher,[3] Haichuan Liu,[3] Steve Hall,[3] Evelin D. Szakal,[3] H. Ewa Witkowska,[3] Adam A. Arkin,[1] Gareth P. Butland,[1] Swapnil Chhabra,[1] John Tainer,[1] Greg Hura,[1] Mark Biggin,[1] Kenneth Downing,[1] Robert Glaeser,[1] Manfred Auer,[1] Gary Siuzdak,[2] Adam M. Deutschbauer,[1] Paul Adams,[1] and **Trent Northen**[1]* (trnorthen@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]The Scripps Center for Mass Spectrometry, La Jolla, Calif.; and [3]UCSF Sandler-Moore Mass Spectrometry Core Facility, University of California, San Francisco

http://enigma.lbl.gov/

Project Goals: Microorganism-based approaches to address DOE mission goals in remediation, carbon sequestration and energy production will require quantitative understanding of biological complexity at multiple scales—from molecular networks of individual species to the dynamic inter-species interactions within the communities in which they reside. The broad goals of ENIGMA are to understand, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides with sufficient detail to predictively model interactions within microbial and community processes that drive complex geochemistry in key environments. In doing so we expect to define biological principles governing selection of microbial community function and composition in given environments.

To ascertain the key processes contributed by an organism to a community and to characterize that process within the organism in molecular detail, it is necessary not only to identify the components of the pathways involved, but to also understand their functional interactions with other processes in the organism of interest. The Biotechnology Component of ENIGMA is developing and applying a suite of technological approaches, from genetics, protein abundance, structure, localization, and metabolism to enable systems-level insights into microbial activity. We have established a flexible experimental pipeline in metal-reducing and sulfate-reducing bacteria (SRB) for (1) high-throughput strain/construct generation, (2) evidence-based annotation of gene function using mutagenesis and extensive phenotyping, (3) evidence-based annotation of transcripts using tiling microarrays and RNAseq (4) protein and protein complex isolation (TEM and x-ray techniques), (5) mass spectrometric based proteomic analysis, (6) Protein and protein complex structural analysis, (7) mass spectrometry based metabolomics analysis (GC-TOF, LC-MS/MS, NIMS), and (8) high resolution imaging (FIB/SEM, SBF/SEM, PALM, STORM). Further technology development will enable us to apply our approaches to environmental isolates rapidly and cost effectively. Additionally, we are exploring the integration of diverse data types including metabolomics and high-throughput genetics to elucidate gene function. Long term, these diverse data will form the foundation for predictive models for a number of key microorganisms from a single environment thereby providing a rich resource for assessing ecological questions relevant to microbial community structure and function.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

194

ENIGMA Biotechnology: Metabolic Profiling of Bacterial Mutant Libraries to Link Metabolite Utilization to Genotype

Richard Baran* (RBaran@lbl.gov), Benjamin Bowen, Morgan Price, Adam A. Arkin, Adam Deutschbauer, and **Trent Northen**

Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://enigma.lbl.gov/

Project Goals: Microorganism-based approaches to address DOE mission goals in remediation, carbon sequestration and energy production will require quantitative understanding of biological complexity at multiple scales—from molecular networks of individual species to the dynamic inter-species interactions within the communities in which they reside. The broad goals of ENIGMA are to understand, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides with sufficient detail to predictively model interactions within microbial and community processes that drive complex geochemistry in key environments. In doing so we expect to define biological principles governing selection of microbial community function and composition in given environments.

The discrepancy between the pace of sequencing and functional characterization of genomes has been recognized as one of the major challenges in microbial genomics. The ENIGMA program at LBNL is addressing this by integrating high-throughput genetics and mutant libraries with metabolic observations to establish direct functional associations between genes and metabolic processes. Mass spectrometry (MS) based metabolomics allows the profiling of metabolites in complex biological samples with high sensitivity and is well suited for interrogating the metabolic capabilities of microorganisms. We have used untargeted metabolomics to identify unexpected and novel metabolites as well as to map the uptake and release of a broad range of metabolites by different microorganisms. Large, genome-wide libraries of mutant strains, developed for multiple ENIGMA microorganisms, enable investigations into the role of specific genes in various metabolic processes using high-throughput metabolomics. However, metabolite

profiling of complex samples is usually performed using time-consuming chromatography to separate the metabolites prior MS analysis. This makes it largely incompatible with screening large mutant libraries. To meet this challenge, we developed a workflow combining untargeted and high-throughput metabolomics to identify genes of *Escherichia coli* and *Shewanella oneidensis* MR-1 related to the utilization of specific metabolites. The ability to uptake specific metabolites from different complex media was analyzed using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). A subset of metabolites found to be utilized were supplemented to minimal media. A limited complexity of the media facilitated high-throughput analysis of spent media extracts using MS. A total of 8000 mutant strains were screened for defects in metabolite utilization. The presence of one of the tested metabolites in the spent media extracts of specific mutants directly links genes to a metabolic defect. Intracellular metabolites of these mutants were then profiled by LC-MS to identify potential accumulation of intermediates related to the utilization of specific metabolites. Using this approach, we identified genes of known function as well as putative transport proteins and enzymes with previously ambiguous annotations.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

195

ENIGMA – Biotechnology: High-Throughput Chemogenomic Fitness Profiling of ENIGMA Relevant Microbes Using a Next Generation Barcode Sequencing Assay

Adam M. Deutschbauer,[1] Jennifer He,[1] Morgan Price,[1] Jennifer V. Kuehl,[1] Kelly Wetmore,[1] Nancy L. Liu,[1] Sonia Reveco,[1] **Swapnil R. Chhabra**,[1] Grant M. Zane,[2] Tom R. Juba,[2] **Judy D. Wall**,[2] **John-Marc Chandonia**,[1] **Trent R. Northen**,[1] **Adam P. Arkin**,[1] and **Gareth P. Butland**[1]* (GPButland@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]University of Missouri, Columbia

Project Goals: ENIGMA's goal is to understand, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides. Environmental change or human intervention will alter the chemical environment in the subsoil, which in turn affects which species predominate as well as microbial physiology. It is therefore critical to understand how microbes respond to environmental changes which may invoke a cellular stress response and inhibit growth or alternatively provide an opportunity to thrive within a community. A comprehensive understanding of the pathways impacted by specific stresses or changes in environment, together with data from other ENIGMA groups detailing changes in community composition as a whole under similar conditions will allow us to link the viability of specific organisms of interest to their ability to contribute or perform specific functions for or within the community.

To ascertain the key processes contributed by an organism to a community and to characterize that process within the organism in molecular detail, it is necessary not only to identify the components of the pathways involved, but to also understand their functional interactions with other processes in the organism of interest. To address these questions, we have utilized recently created comprehensive genome wide collections of single mutant knockout strains and performed chemogenomic fitness profiling to assess the change in viability (growth fitness) of thousands of mutant strains under a variety of growth conditions (i.e. presence of stress agent, change in growth medium, etc.) in a high throughput format. Chemogenomic interactions are identified when a specific mutant or collection of mutant strains have a statistically significant defect or growth advantage under a specific condition when compared to other mutants, and are a powerful indicator of functionally related genes which may be present in the same or related pathways and processes.

Previously used methods to assay the fitness of mutant strains in model systems have relied on the imaging of colony growth on agar plates. This method is not an option for a potentially diverse range of microbes of interest to ENIGMA which may not form morphologically uniform colonies or may excrete chromophoric substances into the media, precluding accurate imaging. However, due to the successful development and application of a TagModule based approach to create mutant strains of *Shewanella oneidensis* MR-1, *Desulfovibrio alaskensis* G20 and *Desulfovibrio vulgaris* Hildenborough with dual unique 20nt molecular barcodes, we are leveraging these strain collections to perform pooled fitness assays, containing thousands of mutants in a single sample. Microarray hybridization or next generation sequencing can be used as a readout for the change in abundance of specific mutant barcodes within the pool under specific stress conditions.

To this end, we have recently designed and implemented an assay, based on pioneering work in yeast, which utilizes the sequencing of unique molecular barcodes on an Illumina platform as a quantitative readout for mutant strain fitness. This assay, as with a microarray readout, is species independent but will allow for massively increased throughput and dynamic range compared to existing approaches at a lower per experiment cost. Initial development of this assay utilized a collection of *S. oneidensis* mutants, each containing a different pair of molecular barcodes, due to the large quantity of microarray based chemogenomic fitness data available from this collection for comparative analysis. Preliminary assay data comprised over 173 million usable barcode reads for a single Illumina lane. This massive sequencing capacity has allowed us to utilize experimental indexes to analyze multiple conditions on the same Illumina lane. Using an initial 18-index multiplex, for an ~4000 mutant complexity pool, we were able to obtain an average of ~2000 barcode

reads for each mutant present (typical range 700 to 3000). These data and previous work in yeast suggests that at minimum, 48-index multiplexing (24-48 stress conditions per lane) is feasible. As our Illumina library preparation protocol optimization is complete, we will apply this technique to perform chemogenomic fitness profiling of single gene mutants of *Desulfovibrio alaskensis* G20 and *Desulfovibrio vulgaris* Hildenborough to highlight the hypersensitivity and resistance of strains to stress conditions, chemical agents and diverse metabolic conditions, thereby providing critical insight into gene function and pathway composition.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

196

ENIGMA Biotechnology: Identification of Differentially Expressed Metabolites in *Desulfovibrio* and Communities Using Meta-Analysis and the METLIN Database

Richard Baran,[1] Kevin Cho, [2] Benjamin Bowen,[1] Paul Adams,[1] Adam A. Arkin,[1] Adam M. Deutschbauer,[1] Linh Hoang,[2] Ewa Kalisiak,[2] Nitin S. Baliga,[3] Zhengjiang Zhu,[2] Ralf Tautenhahn,[2] Trent R. Northen,[1] and **Gary Siuzdak**[2]* (siuzdak@scripps.edu)

[1]Lawrence Berkeley National Laboratory, Berkeley Calif.; [2]The Scripps Center for Mass Spectrometry, La Jolla Calif.; and [3]Institute for Systems Biology, Seattle Wash.

http://enigma.lbl.gov
http://masspec.scripps.edu/

Project Goals: Our objectives are to examine microorganisms and their communities and measure both their chemical input and output as a read out of specific biochemical activity. This effort will build on our previous experience of complex microbial systems by examining how bacteria communities rely on symbiotic relationships for survival and reproductive success, and how these relationships consequently affect their biochemical capabilities.

Metabolomics has emerged as a powerful tool to interrogate cellular biochemistry at the systems level by tracking alterations in the levels of small molecules. The success of metabolomics over the past decade has largely relied on advances in mass spectrometry instrumentation, coupled with developments in bioinformatic tools such as XCMS-Online, it has now become relatively routine to comprehensively compare the levels of metabolite peaks. To facilitate metabolite identification within the ENIGMA program we have developed a freely accessible metabolite database called METLIN (http://metlin.scripps.edu) which contains tandem mass spectral data from thousands of metabolites. This repository allow investigators to compare MS2 data from their research samples to MS2 data from model compounds catalogued in the database and thereby improve the speed, efficiency, and cost effectiveness of untargeted studies. One approach to define cellular dynamics with respect to alterations in of small molecules has been to consider metabolic flux. While flux measurements have proven effective for model organisms, acquiring multiple time points at appropriate temporal intervals for many sample types is challenging. As an alternative, meta-analysis provides another strategy for delineating metabolic cause and effect perturbations. This combination of metabolomic data from multiple genotypes or environmental conditions enables the association of specific changes in small molecules with unique phenotypic alterations. We recently developed metabolomic software called metaXCMS to automate these types of higher order comparisons. Here we discuss the utility of metaXCMS for analyzing proteomic datasets and highlight the biological value of combining meta-results from both metabolomic and proteomic analyses. The combined meta-analysis has the potential to facilitate efforts in functional genomics and the identification of metabolic disruptions related to specific phenotypes. In particular, we present this approach to characterize strains of *Desulfovibrio alaskensis* and microbial communities. The output of these experiments include the identification of novel endogenous metabolites as well as proteins uniquely associated to a range of pathways including glycolysis, the citric acid cycle, the urea cycle, and select amino acid metabolism.

Informatics Development. Our **XCMS/Metlin** platform consists of a continuously evolving technology that has become the world standard for metabolomics data analysis. **Metlin (http://metlin.scripps.edu/)** the largest tandem mass spectrometry metabolite database, has had over 10 million hits containing over 45,000 structures, and MS/MS data on over 5000 metabolites. We are currently developing Metlin with the addition of ENIGMA bacterial metabolites as well as extending its functionality within **XCMS**. XCMS is the most cited metabolomics software tool with over 50,000 downloads. **XCMS ONLINE** is being developed (**https://xcmsonline.scripps.edu/**) to facilitate the analysis of microbial studies and already has over 1500 users in 36 countries. **XCMS/Metlin** is being designed with an easy to use command driven interface with direct connection to **Metlin** for molecule identification. XCMS ONLINE is also being developed for bacterial analysis to allow for second-order ("meta") analysis (**Figure 1**). MetaXCMS (Anal. Chem. 2011, **Nature Protocols** in press) allows for data reduction when analyzing multiple sample groups to provide information on key proteins/metabolites related to a phenotype (such as *Desulfovibrio desulfuricans* metal reduction).

Proteomics/Metabolomics Analytical Development. Our mass spectrometry based technology developments include quantitative **QqQ MRM** metabolomics methods, **LC/MS/MS** shotgun high throughput protein analysis (**Nature 2010**) as well as novel **CESI-MS** (capillary electrophoresis mass spectrometry). In particular, we are applying these methods to genetically characterized strains of *Desulfovibrio desulfuricans* including the G100 wild type and the G20 mutant. The output of these experiments include the identification of novel endogenous metabolites (agmatine, cytosine, and 7-hexadecenoic acid) and 66 proteins uniquely

present in the G100 wild type organism. These approaches are being developed for ENIGMA are allowing us to examine glycolysis, the citric acid cycle, the urea cycle, and select amino acid metabolism among other pathways.

Community Flux by Pulse labeling with a Biological Event. Although very biologically informative, isotopic pulse labeling flux experiments are extremely difficult in the context of mass spectrometry. Especially when examining systems that are largely unknown, stable isotopes cause a shift in mass readout that complicates databases and limits analysis by current bioinformatic software. Therefore, as an alternative to stable isotopes, we are pulse labeling with a biological event to examine temporal changes with respect to that event. These experiments are being performed with different populations of the community where the primary purpose is to identify the key biochemical perturbations that occur as a response to that particular event.

References

1. Patti G.J., Yanes O., Siuzdak G. Metabolomics: leading from behind *Nature Reviews, in press*, 2012.
2. Patti G.J., Yanes O., Shriver L., Courade J-P., Tautenhahn R., Manchester M., Siuzdak G. Metabolomics Implicates Altered Sphingolipid Metabolites *Nature Chemical Biology, in press*, 2012.
3. Patti G.J., Tautenhahn R., Siuzdak G. Meta-Analysis of Untargeted Metabolomic Data from Multiple Profiling Experiments *Nature Protocols, in press*, 2012.
4. Patti G.J, Tautenhahn R., Fonslow B.R., Cho Y., Deutschbauer A., Arkin A., Northen T., Siuzdak G. Meta-Analysis of Global Metabolomics and Proteomics Data to Link Alterations with Phenotype *Spectroscopy, in press (cover article)*, 2012.
5. Tautenhahn R., Patti G., Kalisiak E., Miyamoto T., Schmidt M., Fan Y.L., McBee J., Baliga N.S., Siuzdak G. metaXCMS: Second-Order Analysis of Untargeted Metabolomics Data *Analytical Chemistry*, 2011, 83, 696-700.
6. Yanes O., Tautenhahn R., Patti G., Siuzdak G. Expanding Coverage of the Metabolome for Global Metabolite Profiling *Analytical Chemistry*, 2011, 83, 2152-2161.
7. Greving M.P., Patti G.J., Siuzdak G. Nanostructure-Initiator Mass Spectrometry (NIMS) Metabolite Analysis and Imaging *Analytical Chemistry*, 2011, 83(1), 2-7.
8. Long J.Z., Cisar J.S., Miliken D., Niessen S., Wang C., Trauger S.A., Siuzdak G., Cravatt B.F. Metabolomics annotates ABHD3 as a regulator of medium-chain phospholipids *Nature Chemical Biology, in press*, 2011.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

197

ENIGMA Biotechnology: Membrane Protein Complexes—Their Roles in *Desulfovibrio vulgaris* Stress Response and in the Establishment and Maintenance of Communities

Peter J. Walian[1]* (PJWalian@lbl.gov), Simon Allen,[2] Max Shatsky,[1] Lucy Zeng,[1] Evelin D. Szakal,[2] Haichuan Liu,[2] Bonita Lam,[1] Jil T. Geller,[1] Kristina L. Hillesland,[3] Steven C. Hall,[2] Susan J. Fisher,[1,2] **Matthew W. Fields**,[4] **David A. Stahl**,[3] **Terry C. Hazen**,[5] **Steven E. Brenner**,[6] **Adam M. Deutschbauer**,[1] **Trent R. Northen**,[1] **John-Marc Chandonia**,[1,6] **H. Ewa Witkowska**,[2] **Mark D. Biggin**,[1] **Bing K. Jap**,[1] and **Paul D. Adams**[1]

[1]Lawrence Berkeley National Laboratory; [2]University of California, San Francisco; [3]University of Washington; [4]Montana State University; [5]University of Tennessee, Knoxville; and [6]University of California, Berkeley

http://enigma.lbl.gov

Project Goals: Key aims of this project are—to develop a system for the high-throughput isolation and identification of membrane protein complexes, optimizing this process for effectiveness across a range of sample types including planktonic cultures and biofilms; to apply this system in the study of DOE relevant microbes such as *Desulfovibrio vulgaris* in order to detect and characterize changes in their membrane protein complexes brought about by environmental stressors, and through the role of these proteins in the establishment and maintenance of communities.

A central goal of the ENIGMA consortium is to develop robust molecular-level models capable of predicting how target microbes respond to a range of environmental conditions. In support of this goal, our interests within ENIGMA have centered on the dynamic role of membrane protein complexes in this process. Cell membranes represent the "front-line" of cellular defense and the interface between a cell and its environment. Significant changes in response to environmental conditions are expected to take place through the proteins situated within these membranes. Membrane protein-associated changes may occur in the form of abundance level, protein-protein interactions, post-translational modifications and even mutations. To understand some of the earliest and perhaps most critical responses to stress, characterization of these changes on a molecular level is needed.

The study of membrane proteins presents a major challenge in protein biochemistry; to address this we have developed a unique high-throughput process for the isolation and identification of untagged membrane protein complexes that features mild, but effective, detergent solubilization, liquid chromatography and native electrophoresis methods. We have been applying this system in two main areas of

investigation, one of which has been our work on developing a *D. vulgaris* membrane protein complex database covering standard and stressed growth conditions, and the second to characterize the roles of selected membrane proteins in the establishment and maintenance of communities.

Our study of membrane protein complexes in the outer-membrane of *D. vulgaris* grown under standard conditions is complete and we are at an advanced stage with the inner-membrane component. An interactome of proteins identified in *D. vulgaris* outer-membrane preparations is in the final stages of refinement. These preparations have yielded 69 outer-membrane protein identifications (which is over 80% of the number expected); 90% of these proteins were found to be in complexes. The most prevalent categories of proteins detected were the lipoproteins, and proteins with non-specific annotations (hypothetical and conserved hypothetical). This compendium of *D. vulgaris* outer-membrane protein complexes will serve as an essential reference for the detection and characterization of environment-driven changes in these proteins. Processing of outer-membrane proteins from stressed *D. vulgaris* cultures (including growth to stationary phase, and growth under elevated levels of nitrate or NaCl) has recently been completed. Initial analysis of stress-associated changes in outer-membrane protein abundance suggests that for many proteins there are significant differences between these changes and the changes in expression level inferred from mRNA experiments. Efforts on the preparation of the inner-membrane protein interactome and completion of the analysis of stress-induced changes occurring in the outer-membrane proteins of *D. vulgaris* are on-going.

In addition to our studies on large-volume planktonic monocultures, we have refined the methods employed in the pipeline so that they may be used to process samples derived from more native-like sources. Recent improvements made in pipeline sensitivity and resolution are now allowing us to work successfully with relatively small cell samples such as biofilms. We anticipate that through additional optimization of the system, we will be able to process yet smaller samples, not only cultured in the lab but obtained directly from field sites. To assess the potential for discovery from such sources, we have begun pilot studies on biofilm samples. Early results contain evidence of protein changes occurring during the transition from stationary phase to biofilms, suggesting that this will be a productive direction for future studies of microbial communities. Recent work by the Stahl group on adaptive evolution experiments with co-cultures of *Methanococcus maripaludis* and *D. vulgaris,* has identified mutations of soluble and membrane protein genes likely to be important in establishing the syntrophic mutualism between these species. These results suggest a large influence of acquired membrane protein mutations leading to improved growth rates within this co-culture community. We have begun to process membranes from clonal isolates to characterize changes in the *D. vulgaris* membrane protein population. Interestingly, the most abundant protein of the *D. vulgaris* outer-membrane (DVU_0799) is also the most consistently mutated protein in these experiments. Therefore, a key goal of ours will be to purify and functionally characterize those mutated membrane proteins found to play a role in facilitating improved rates of growth.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

198

ENIGMA Biotechnology: Biofilm Imaging: From Protein Complexes to Intact Microbial Communities

M. Auer,[1] Amita Gorur[1]* (agorur@lbl.gov), P. Arbeleaz,[2] N. Baliga,[3] D.A. Ball,[1] M. Biggin,[1] J.M. Chandonia,[1] S. Chhabra,[1] R. Csencsits,[1] K.H. Downing,[1] M.W. Fields,[4] J.T. Geller,[1] R. Glaeser,[1], T.C. Hazen,[1] G.L. Hura,[1] T. Juba,[5] B. Lam,[1] C.M. Leung,[1] J. Liphardt,[2] J. Malik,[2] J.P. Remis,[1] S. Reveco,[1] J. Tainer,[1] A. Tauscher,[1] J. Wall,[5] A. Deutschbauer,[1] T. Northen,[1] A. Arkin,[1] and **P.D. Adams**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley Calif.; [2]University of California, Berkeley; [3]Institute for Systems Biology, Seattle, Wash.; [4]Montana State University, Bozeman; and [5]University of Missouri, Columbia

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project aims to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. The ENIGMA Biotechnology Component focuses its efforts on providing cross-cutting technologies that will support all other ENIGMA components with a particular focus on biological imaging at different scales of bacteria and microbial communities. These data will help to develop models of microbial community activity and principles of community organization in an effort to predict the role that microbial species and their interactions play in the dynamics of geochemical transformations in a changing environment.

Microbial physiology is inherently a multiscale biological process that coordinates complex processes such as extracellular metal reduction and response to environmental stresses and competing species. Bacteria often assemble into sustainable communities that allow individual bacteria to coordinate their respective behavior and thus optimizing the efficiency of biological processes, which may enhance the chances for species survival.

ENIGMA is addressing the complexity of multiscale spatiotemporal biofilm organization through a combination of expertise in traditional structural biology and modern multimodal imaging. SAXS (Rambo *et al.* 2010) and single particle cryo-EM are proven technologies to determine protein complex stoichiometry and shape, allowing the fitting of high-resolution structures into the intermediate resolution density envelope (Han *et al.* 2009). Cryo-electron tomography of bacterial whole mount samples can detect intra- and extracelluar specializations e.g. those important

for metal reduction. Cryo-EM analysis is complemented by widefield 2D section TEM and advanced 3D SEM imaging approaches (FIB/SEM and SBF/SEM) of cryo-preserved, freeze substituted and resin-embedded samples. With these novel EM imaging approaches, we have begun to examine large areas and volumes of biofilms in DvH and other soil bacteria. We have found outer membrane vesicles, vesicle chains and cell-cell connections (Palsdottir *et al.* 2009, Remis *et al*. 2010, Remis *et al.* submitted), as well as compartmentalization of metal precipitation (Auer, unpublished observation). These observations suggest a an intricate set of interactions and possibly coordination of function between community members. X-ray and EM-based imaging approaches are complemented by tag-based labeling of proteins both at the light and electron microscopy level, and allow the study of cell-to-cell variations in protein abundance and protein localization (Chabra *et al.* 2010). Advanced optical super-resolution imaging methods (including PALM and STORM) allow high precision localization and counting (Betzig *et al* 2006). Further integration of small molecule mass spectrometry imaging, while at a somewhat larger size scale, promises to link structural observation and protein localization with metabolic activity of biofilm regions. Through the integrated application of these imaging modalities ENIGMA is deconstructing a mechanistic understanding of biofilm function.

Acknowledgement: This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

1. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science 2006 313 (5793): 1642–1645
2. Chhabra SR, Butland G, Elias DA, Chandonia JM, Fok OY, Juba TR, Gorur A, Allen S, Leung CM, Keller KL, Reveco S, Zane GM, Semkiw E, Prathapam R, Gold B, Singer M, Ouellet M, Szakal ED, Jorgens D, Price MN, Witkowska HE, Beller HR, Arkin AP, Hazen TC, Biggin MD, Auer M, Wall JD, Keasling JD. Generalized schemes for high-throughput manipulation of the *Desulfovibrio vulgaris* genome. Appl Environ Microbiol. 2011 Nov;77(21):7595-604
3. Han BG, Dong M, Liu H, Camp L, Geller J, Singer M, Hazen TC, Choi M, Witkowska HE, Ball DA, Typke D, Downing KH, Shatsky M, Brenner SE, Chandonia JM, Biggin MD, Glaeser RM. Survey of large protein complexes in *D. vulgaris* reveals great structural diversity. Proc Natl Acad Sci U S A. 2009 Sep 29;106(39):16580-5.
4. Palsdottir H, Remis JP, Schaudinn C, O'Toole E, Lux R, Shi W, McDonald KL,Costerton JW, Auer M. Three-dimensional macromolecular organization of cryofixed Myxococcus xanthus biofilms as revealed by electron microscopic tomography. J Bacteriol. 2009 Apr;191(7):2077-82
5. Rambo RP, Tainer JA. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. Curr Opin Struct Biol. 2010 Feb;20(1):128-37.
6. Remis JP, Costerton JW, Auer M. Biofilms: structures that may facilitate cell-cell interactions. ISME J. 2010 Sep;4(9):1085-7
7. Remis JP, Wei W, Gorur A, Allen S, Witkowski HE, Costerton JW, Auer M Bacterial Social Networks: Targeted Delivery of Outer Membrane Proteins via Membrane Vesicle Chains, submitted

199

ENIGMA Computation: An Integrated Framework of Databases and Software for Exploring Genes, Genomes, Proteomes, and Networks in Microbes and Communities

J. Christopher Bare[1]* (Christopher.Bare@systemsbiology.org), Paramvir S. Dehal,[2] Karen Foley,[1] Keith Keller,[2] Morgan N. Price,[2] Wei-Ju Wu,[1] Serdar Turkarslan,[1] Jason Baumohl,[2] Marcin P. Joachimiak,[2] John T. Bates,[2] David J. Reiss,[1] Pavel S. Novichkov,[2] **Adam P. Arkin**,[2] **Nitin S. Baliga**,[1] and **Paul D. Adams**[2]

[1]Institute for Systems Biology, Seattle, Wash. and [2]Lawrence Berkeley National Laboratory, Berkeley Calif.

http://microbesonline.org

Project Goals: Currently, one of the grand challenges in biological discovery is not the production of data but it is the data processing and analysis. Often, algorithms or computational tools exist tailored to the analysis of specific datasets yet they lack modularity to handle new or non-standard data formats, automation for non-expert usage or connectivity with downstream applications. Therefore it is critical to develop biology driven accessible computational frameworks that will promote, support and validate biological discovery. In order to tackle this grand challenge, the goals of this project is to i) integrate existing tools and ongoing efforts for algorithm, software and database development within ENIGMA, ii) develop an automated and modular framework for storing, analyzing and visualizing biological data for microbes or communities iii) promote knowledge transfer within ENIGMA and in the scientific community by providing a suite of computational resources.

While current technological advances enabled production of massive amounts of biological data, the real challenge is to connect these data sources to biological discovery. Here we describe our efforts to develop a modular and integrated framework for storage, automated analysis and visualization of biological datasets. Using Microbes Online as a source of high-quality curated data, we infer regulatory networks for several organisms. Feeding gene expression data and experimental metadata through the cMonkey and Inferelator, network inference algorithms produce, for a given organism, a set of co-expressed and putatively co-regulated modules. These networks are archived in the Network Portal, which provides a web interface for searching, visualizing and annotating the networks. The Network Portal will be tightly inte-

grated with Microbes Online, which already provides a rich set of features for comparative microbial genomics as well as RegPrecise for curation of regulatory inferences. Additional experimental validation of network inferences can be performed using the MicrobesOnline Fitness Experiment data. Automation of the network inference pipeline will enable us to capture phylogenetic diversity to drive comparative studies and availability of high quality fitness data will help us to asses and curate network inferences. In addition, the Gaggle integration framework provides interoperability with several popular bioinformatics tools including Cytoscape, MeV, and R. Firegoose provides Gaggle connectivity to web resources such as KEGG, DAVID and EMBL STRING. This suite of integrated analysis and visualization tools provides a powerful and easily extended environment for the study of the regulatory systems of microbial biology.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

200

ENIGMA Computation: Inferring Correlation Networks from Genomic Survey Data

Yonatan Friedman and **Eric J. Alm*** (ejalm@mit.edu)

Computational and Systems Biology Initiative, Massachusetts Institute of Technology, Cambridge

High-throughput sequencing based techniques, such as 16S profiling, have the potential to elucidate the complex inner workings of natural communities - be they the world's oceans or the human gut. A key step in exploring such data is the identification of dependencies between members of these communities, which is commonly achieved by means of correlation analysis.

However, it has been known since the days of Karl Pearson that the analysis of the type of data produced by such techniques (referred to as compositional data) can produce unreliable results since the observed data take the form of relative fractions of genes or species, rather than their absolute abundances.

Inferring correlation networks between genes or species is a common goal in microbial ecology. However, such networks are typically derived from genomic survey data, such as those obtained from 16S sequencing, which are subject to underappreciated mathematical difficulties that can undermine standard data analysis techniques. We show that these effects can lead to erroneous correlations despite the statistical significance of the associations. To overcome these difficulties, we developed SparCC; a novel procedure, tailored to the properties of genomic survey data that allow inference of correlations between genes or species.

201

ENIGMA Computation: Multiscale Modeling of Gene Regulatory Networks Across Evolutionary Timescales by Integrating "Top-Down" and "Bottom-Up" Approaches

Pavel Novichkov,[1]* David J. Reiss[2]* (dreiss@systemsbiology.org), Marcin Joachimiak,[1] Alexey Kazakov,[1] Serdar Turkaslan,[2] **Inna Dubchak**,[1] **Adam P. Arkin**,[1] and **Nitin S. Baliga**[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]Institute for Systems Biology, Seattle, Wash.

http://enigma.lbl.gov/ENIGMA_InformaticsKnowledgebase.html
http://baliga.systemsbiology.net/enigma

Project Goals: Environmental conditions and their associated cellular states dynamically shape regulatory responses. As the environment changes, the solution space for regulation of the same gene (or set of genes) moves between different states depending on the combinatorial interplay between transcription factors, their ligands, and binding sites. Such responses are encoded in gene regulatory networks by specific cis-regulatory signals encoded in the genome. We have developed a novel strategy for reverse engineering microbial gene regulatory networks that times context-specific system-scale regulatory phenomena directly to mechanism at single nucleotide-level resolution.

As a first step toward this goal we have integrated the data-driven and genomic-driven regulon discovery approaches developed within the ENIGMA team. The user can start from an automatic gene regulatory network inference conducted with the ***cMonkey*** integrated biclustering algorithm. ***cMonkey*** generates biclusters (predicting putative regulons with corresponding putative condition-specific cis-regulatory motifs) which can then be directly input to ***RegPredict*** for detailed comparative genomics analysis in order to verify evolutionary conservation of transcription factor binding sites (TFBS) and to refine the regulon content. ***RegPredict*** automatically suggests the most appropriate set of closely related genomes required for comparative analysis. Finally, the refined regulon can be submitted to the ***RegPrecise*** database, a community repository of manually curated regulons.

We have used this strategy to build transcriptional regulatory network models for ENIGMA keystone organisms *D. vulgaris* Hildenborough and *M. maripaludis*. The workflow was tested on several regulons (Rex, Fur, FliA, PerR, SahR) from *D. vulgaris* Hildenborough. For example, ***cMonkey*** predicted the Fur regulon and its corresponding TFBS, which was then refined and 19 additional gene members added using the comparative genomics approach of ***RegPredict***. Based on this model, we are also developing a framework for automated network inference plugged into a Network Portal, which includes a suite of analysis and visualization tools.

The data-driven systems approach has proven powerful for automatic reconstruction of mid- to large-size regulons. However, it is more challenging to identify small regulons due to their small size, a property that limits both gene expression data comparisons and sequence motif detection. Our novel method enables us to tackle the small regulon portion of the regulatory network reconstruction problem. The method uses a new biclustering engine along with heuristics designed from properties of known regulons to identify small regulon patterns in phylogenetic profiles.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

202

ENIGMA Computation: Evolutionary and Experimental Evidence-Based Functional Annotation of Genes

Morgan N. Price,[1] **Adam P. Arkin**,[1,2] Ben Bowen,[3] Barbara E. Engelhardt,[4,5] Ameet Talkwalkar,[4] Jeffrey M. Yunes,[6] Michael L. Souza,[7] Gaurav Pandey,[8,9] Susanna Repo,[8] Michael I. Jordan,[4,10] **Steven E. Brenner**,[1,2,6,7,8]* (brenner@compbio.berkeley.edu), Nitin Baliga,[11] and Paul D. Adams[1]

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]Department of Bioengineering, University of California, Berkeley; [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; [4]Department of Electrical Engineering and Computer Science, University of California, Berkeley; [5]Currently at Biostatistics and Bioinformatics Department, Duke University, Durham, N.C.; [6]Joint Graduate Group in Bioengineering, University of California, Berkeley and San Francisco; [7]Biophysics Graduate Group, University of California, Berkeley; [8]Department of Plant and Microbial Biology, University of California, Berkeley; [9]Currently at Mount Sinai School of Medicine, New York; [10]Department of Statistics, University of California, Berkeley; and [11]Institute for Systems Biology, Seattle, Wash.

http://enigma.lbl.gov
http://compbio.berkeley.edu

Project Goals: The ENIGMA SFA aims to understand the architecture of microbial communities from a molecular level, which requires understanding in detail the molecular biology of key organisms. Although sequencing the genomes of these organisms is now straightforward, determining the molecular function of genes remains a challenge. However, many genome sequences are now available, and rich genome-wide functional data is becoming available outside of traditional model organisms. Therefore, we are developing improved tools for using evolutionary comparisons and functional-genomic data to predict the molecular function of proteins.

Phylogenetic analysis has been employed to infer the molecular function of a target gene by finding a function that is consistent with the evolutionary history of the gene. Over the past decade, this has been recognized as a highly accurate approach, but its manual application requires laborious effort by a domain expert. We have developed the SIFTER method, which automates phylogenetic-based function annotation by finding the most likely assignments of functions to proteins given a phylogenetic tree, model of evolution, and known functions. SIFTER uses a Bayesian graphical model framework to propagate molecular functions across the tree in a way that is statistically rigorous and robust. SIFTER explicitly takes account of evidence quality, to account for the variable quality of annotations from different sources.

Benchmarking studies of SIFTER show that it outperforms other widely-used homology-based approaches. Recently, we improved the core SIFTER algorithm, enabling it to run on large and diverse protein families, to work on a genome-scale, and to participate in the Critical Assessment of Function Annotation in 2011. We are extending it to share information between protein families based on gene-gene "association" relationships such as protein-protein interactions, co-expression, co-fitness, genome proximity, or genetic interactions. In doing so, we will be able to incorporate a larger variety of experimental data developed by and applicable to the ENIGMA project than other prediction approaches. We hope that with these enhancements, SIFTER will be the first successful method to statistically incorporate both homology and association data.

Another challenge is to interpret large-scale "fitness" data or knockout mutant phenotypes that are becoming available for diverse microbes due to approaches such as tagged transposon mutagenesis or TnSeq. In a pilot study in *Shewanella oneidensis* MR-1, we were able to confirm many annotations and to revise the annotations of 40 genes or operons, but this required extensive manual curation. To streamline the analysis, we are developing heuristics to find "re-annotatable" proteins to focus the manual curation. We are also automating the comparison of fitness data to metabolic models; in principle, it should be possible to automate much of the manual curation that now goes into producing a high-quality metabolic model.

This work partially conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was partially supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Other support for SIFTER comes from NIH K22 HG00056, NIH R01 GM071749, DOE SciDAC BER KP110201.

203

ENIGMA Computational Core Group: Progress and Directions

David Reiss,[1] Serdar Turkarslan,[1] Jason Baumohl,[3] Paramvir Dehal,[3] Marcin Joachimiak,[3] Keith Keller,[3] Morgan Price,[3] Max Shatsky,[3] Alexey Kazakov,[3] Pavel Novichkov,[3] Yonatan Friedman,[2] Sonia Timberlake,[2] Michael Souza,[4] John-Marc Chandonia,[3] Steven Brenner,[4] Inna Dubchak,[3] Adam P. Arkin,[3] **Eric Alm**,[2] **Nitin Baliga**[1]* (Nitin.Baliga@systemsbiology.org), and Paul D. Adams[3]

[1]Institute for Systems Biology, Seattle, Wash.; [2]Massachusetts Institute of Technology, Boston; [3]Lawrence Berkeley National Laboratory, Berkeley Calif.; and [4]University of California, Berkeley

Project Goals: The ENIGMA Computational Core Group is tasked with data management analysis and dissemination for the ENIGMA Scientific Focus Area. The group develops and maintains widely used computational tools including the MicrobesOnline, RegTransBase, and RegPrecise databases, as well as data integration and visualization tools like the Gaggle, which also support a wide external user base. Analytical efforts have focused on regulatory network prediction (from genome sequence data), network inference (from functional data), gene functional annotation, and statistical approaches for environmental genomics. Efforts in the Computation Core Group integrate all of the projects in ENIGMA. In addition to developing algorithms to find patterns within noisy high throughput systems biology data (e.g. transcriptome structure, peptide atlas, protein-DNA interactions, and biodiversity from NextGen sequencing data), the Computation Core integrates diverse data types to elucidate modular architectures across scales (e.g. community assemblages and regulons) and infers genetic and environmental influences that encode dynamical interrelationships across these modules (e.g. microbial community networks and GRNs).

Using comparative analysis we have tracked the evolutionary history of gene functions to understand how novel functions evolve. One level up, we have used proteomics data, high-resolution genome tiling microarrays, and 5' RNA sequencing to revise genome annotations, discover new genes including ncRNAs, and map dynamically changing operon structures of five model organisms including *Desulfovibrio vulgaris* Hildenborough, *Pyrococcus furiosis*, *Sulfolobus solfataricus*, *Methanococcus maripaludis* and *Halobacterium salinarum* NRC-1. We have developed machine learning algorithms to accurately identify protein interactions at a near-zero false positive rate from noisy data generated using tagless complex purification, TAP purification, and analysis of membrane complexes. Further, we have developed algorithms to analyze and assign significance to protein interaction data from bait pull-down experiments and integrate these data with other systems biology data through associative biclustering in a parallel computing environment. We will "fill-in" missing information in these interaction data using a "Transitive Closure" algorithm and subsequently use "Between Commonality Decomposition" algorithm to discover complexes within these large graphs of protein interactions. To characterize the metabolic activities of proteins and their complexes we are developing algorithms to deconvolute pure mass spectra, estimate chemical formula for m/z values, and fit isotopic fine structure to metabolomics data. We have discovered that in comparison to isotopic pattern fitting methods restricting the chemical formula by these two dimensions actually facilitates unique solutions for chemical formula generators. To understand how microbial functions are regulated we have developed complementary algorithms for reconstructing gene regulatory networks (GRNs). Whereas the network inference algorithms cMonkey and Inferelator developed enable de novo reconstruction of predictive models for GRNs from diverse systems biology data, the RegPrecise and RegPredict framework developed uses evolutionary comparisons of genomes from closely related organisms to reconstruct conserved regulons. We have integrated the two complementary algorithms to rapidly generate comprehensive models for gene regulation of understudied organisms. Our preliminary analyses of these reconstructed GRNs have revealed novel regulatory mechanisms and cis-regulatory motifs, as well as others that are conserved across species. Finally, we are supporting scientific efforts in ENIGMA with data management solutions and by integrating all of the algorithms, software and data into a Knowledgebase. For instance, we have developed the RegPrecise database (http://regprecise.lbl.gov), which represents manually curated sets of regulons laying the basis for automatic annotation of regulatory interactions in closely related species. We are also in the midst of scaling up MicrobesOnline to handle the growing volume of sequence and functional genomics data. Over the last year our efforts have been focused on providing support for additional genomic and functional genomic data types. Similarly, we have developed several visualization tools to help with the exploration of complex systems biology datasets. A case in point is the Gaggle Genome Browser (GGB), which was enhanced with visualizations for plotting peptide detections and protein-DNA binding alongside transcriptome structure, plus the ability to interactively filter by signal intensity or p-value. Finally, we recognize that future advances to computational infrastructure cannot be anticipated and new software will be developed as dictated by scientific needs within ENIGMA and elsewhere. To account for this reality of how software environments evolve we have made advances to the Gaggle and Firegoose framework that enables interoperability and integration of diverse software and databases. Specifically, we have updated the R-goose package which provides connectivity between several Gaggle compliant bioinformatics tools and R; and prototyped a JSON based upgrade to the Gaggle protocol to make this environment extensible and more language neutral than the previous Java-based protocol.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

204

ENIGMA Environmental Overview: Field to Lab to Field

Terry C. Hazen[1,2*] (tchazen@utk.edu), **Matthew W. Fields**,[3] Jizhong Zhou,[4] Joy Van Nostrand,[4] Dwayne A. Elias,[2] Jennifer J. Mosher,[2] Steven D. Brown,[2] Tommy J. Phelps,[2] Mircea Podar,[2] Anthony V. Palumbo,[2] Richard A. Hurt,[2] Anup Singh,[5] **Eric Alm**,[6] Mark B. Smith,[6] Dominique C. Joyner,[1,2] Romy Chakraborty,[7] Boris Faybishenko,[7] Kara DeLeon,[3] Jil Geller,[7] Bonita Lam,[7] Tamas Torok,[7] Julian Fortney,[1,2] Sharon E. Borglin,[7] David Stahl,[8] Judy Wall,[9] **Adam P. Arkin**,[7] and **Paul D. Adams**[7] (PDAdams@lbl.gov)

[1]University of Tennessee, Knoxville; [2]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [3]Montana State University, Bozeman; [4]University of Oklahoma, Norman; [5]Sandia National Laboratories, Livermore, Calif.; [6]Massachusetts Institute of Technology, Boston; [7]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [8]University of Washington, Seattle; and [9]University of Missouri, Columbia

http://enigma.lbl.gov

Project Goals: ENIGMA working hypotheses: 1) Key transects in the environment provide constraints on community composition and activity that are discernable at multiple scales. The uranium/nitrate/pH gradient may provide one such transect and we will examine the communities and activities in different *in situ* (and laboratory simulated) uranium/chromium contaminated environments. 2) These environmental constraints change in time due in part to the structure and function of these communities and causal relationships may be discovered and quantified. 3) Community structure is established and maintained by varying factors that include transient populations, niche diversification, optimized interactions, and resource competition. We will be attempting to dissect these by field perturbation and time series studies and laboratory simulations of the environment. 4) There are important stable communities that achieve the above, which we can dissect at the level of molecular interactions. We will drive toward isolation of key organisms, discovery of key activities and interactions, and dissect cellular networks that control when and how these activities are expressed.

The implications of the above hypotheses suggest: 1) There is a core set of metabolic factors that are the direct effectors of metal-reduction in soil and, while exact microbial species may vary, the molecular network will be relatively constant. This effort includes the identification of the biomolecular determinants of metal reduction in key populations that directly and/or indirectly interact with other populations that have programmed responses to important environmental parameters. 2) There are particular variants of these keystone microbes that adapt them for different metals and different concentrations and these are deployed in planktonic and attached communities over time during the reductive and reoxidative processes following stimulation by substrate amendment. 3) At a particular site there are core, relatively stable sub-communities of microbes whose interspecific interactions are responsible for the stable reduction (and ultimate reoxidation) of metals. There is also a core community structure of necessary functional classes of microbes that form a stable "food web" to exploit the available energy in the environment nearly optimally. For example at the Hanford 100-H study site, microbes from soil samples differed from the corresponding groundwater organisms (even at the phyla level) and were more diverse ($p=0.001$). While many of the populations were observed in both groundwater and surrogate sediments, the respective matrices appeared to enrich for particular OTUs. Results do not indicate a large shift in dominant organisms in soil from pre- to post- injection, and this may be due to the microbes remaining dominant from the first stimulation. However, a prevalence of core genera and rare genera were observed across 34 samples while urban and rural genera were less abundant.

Background and Significance. ENIGMA is planning on shifting the principle field focus from Hanford to Oak Ridge. For the last 12 years ORNL has been characterizing, monitoring, and conducting field experiments at DOE's BER ORNL Field Research site. The focus has been on elucidating the mechanisms and efficacy of bioreduction and bioimmobilization of U, one of the DOE's most common waste site contaminants. This has involved a number of field studies including pump tests, hydrological modeling, characterization of sediment and groundwater and amendments of ethanol, bromide, and nitrate, etc. More than 300 wells have been established and characterized and are available for analysis in a searchable database.

Research and Design

- Overarching Driver: Determine microbial community structure and function in both *in situ* environments and constructed consortia—environment to the laboratory.
- Elucidate structure to function during key biogeochemical transformations—*immobilization of metals*
- Determine key succession events and mechanisms—*stability in the context of geochemical and thermodynamic constraints*
- Parameterize critical microbes at the phylogenetic and functional level in conjunction with key biogeochemical variables that together, impact and control environmental activities of interest (*e.g.*, metal-reduction; N flow; C flow)
- Identify key populations, directly and/or indirectly related to activities of interest.
- *In situ* and laboratory consortia will be used to explicate levels of biological organization from populations to proteins, and
- Models will be developed with various bioinformatic tools (*e.g.*, AdaptML, random matrix theory, multivariate statistics) commonly used for genes and proteins but applied to populations overlaid with geochemical parameters and engineering controls.

The current tasks for the Environmental Core are as follows:

Task 1. Optimize omics protocols for environmental samples.

Task 2. Use existing data from the ORNL FRC database to design efficient field sampling studies that maximize the geochemical diversity of study sites. This strategy is expected to enhance the resolution of associations between microbial communities and key geochemical features. Toward this end, we have developed a Monte Carlo search algorithm to optimize site selection for geochemical diversity. We are currently looking for 100 wells that we can do a metagenome analysis in cooperation with the existing, planned SBR IFRC program at ORNL. *Overarching principle for Field Studies and Field Linked Resources:* With the cooperation of the ORNL IFRC and as a team, establish the hypotheses, experiments, sampling, analyses, and schedule as a test plan, including all protocols, amounts, and responsibilities. Prioritize everything so that time, money, shipping, sample limitations, and unforeseen contingencies will not limit delivery of milestones. This includes 50%, 75%, and 95% evaluations of test plans by the team prior to execution, and fatal-flaw analyses at each step.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

205

ENIGMA Environmental Microbiology: Microbial Community Dynamics in Groundwater and Surrogate Sediments During HRC® Biostimulation of Cr(VI)-Reduction

K.B. De Leon,[1,2] B.D. Ramsay,[2] D.R. Newcomer,[3] B. Faybishenko,[4] T.C. Hazen,[5,6] M.W. Fields[1,2]* (matthew.fields@erc.montana.edu), and **Paul D. Adams**[4] (PDAdams@lbl.gov)

[1]Department of Microbiology, [2]Center for Biofilm Engineering, Montana State University; [3]Pacific Northwest National Laboratory; [4]Lawrence Berkeley National Laboratory; [5]University of Tennessee; and [6]Oak Ridge National Laboratory

http://enigma.lbl.gov

Project Goals: The elucidation of bacterial community dynamics for both groundwater and sediment-associated communities over time and space during biostimulation for chromate reduction in relation to geochemical variables.

The Hanford 100-H site is a chromium-contaminated site that has been designated by the Department of Energy Environmental Management as a field study site for *in situ* chromium reduction. In August 2004, the first injection of hydrogen release compound (HRC®) resulted in an increase of microorganisms and a reduction of soluble chromium(VI) to insoluble chromium(III). Little is understood about the microbial community composition and dynamics during stimulation. The aim of this study is to compare microbial communities of groundwater and soil samples across time and space during a second injection of HRC®. A second injection occurred November 2008 and geochemical data collected throughout the study showed an overall decrease in nitrate, sulfate, and chromium(VI). Spatial and temporal water and soil samples (n=34) were collected pre-and post-injection from four wells at the field site. Soil columns constructed from stainless steel mesh were lined with nylon mesh and filled with Hanford soils from the 100-H site. The soil columns were used to represent not only the microbes flowing through the soil via groundwater, but the microbes that require a matrix in order to grow. DNA was extracted from each of the samples and SSU rDNA gene fragments was sequenced via multiplex pyrosequencing. Sequences were refined by length, primer errors, and Ns, and sequences with a high percentage of low Phred quality score values were removed. Python scripts were developed to filter the pyrotag data with respect to quality scores, and the filtering technique was validated with environmental samples. Soil samples differed from the corresponding groundwater (even at the phyla level) and were more diverse ($p=0.001$). While many of the populations were observed in both groundwater and surrogate sediments, the respective matrices appeared to enrich for particular OTUs. Predominant populations for the sediments were *Psuedomonas*, *Acidovorax*, *Clostridium*, *Aquaspirillum*, *Methylibium*, *Anaeromyxobacter* while predominant populations for groundwater were *Psuedomonas*, *Pleomorphomonas*, *Ramlibacter*, *Arthrobacter*, and *Herbaspirillum*. Genera observed only in the sediment included *Marinomonas* while genera observed only in the groundwater included *Desulfonauticus*, *Desulfomicrobium*, and *Syntrophobacter*. Results do not indicate a large shift in dominant organisms in soil from pre- to post- injection, and this may be due to the organisms remaining dominant from the first stimulation. However, a prevalence of core genera and rare genera were observed across 34 samples while urban and rural genera were less abundant. There was a shift from *Acidovorax* to *Aquaspirillum* from upstream (non-stimulated) to downstream soil both pre- and post-injection. Surrogate soil samples indicated similar changes in the soil community in the injection (Well 45) and downstream (Well 41) wells across time, while water samples seem to indicate more of a pre- and post-injection grouping instead of gradual changes across time. Furthermore, while post-injection soil samples indicate a continuing dominance of *Aquaspirillum*, corresponding water samples indicate *Pseudomonas* as a dominant genus. For each well, HRC® injection resulted in increased diversity, but the greatest changes during stimulation occurred in the populations of mid-dominance either between wells or across time. These organisms could be important to consider as possible indicator species in future work.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

206

ENIGMA Environmental: Succession of Hanford Groundwater Microbial Communities During Lactate Amendment and Electron-Acceptor Limitation

Dwayne A. Elias[1]* (eliasda@ornl.gov), Jennifer J. Mosher,[1] Tommy J. Phelps,[1] Mircea Podar,[1] Richard A. Hurt,[1] James H. Campbell,[1] Meghan M. Drake,[1] James G. Moberly,[1] Christopher W. Schadt,[1] Steven D. Brown,[1] Terry C. Hazen,[2] Adam P. Arkin,[3] Anthony V. Palumbo,[1] Boris A. Faybishenko,[3] and **Paul D. Adams**[3]

[1]Oak Ridge National Laboratory, Oak Ridge, Tenn.; [2]University of Tennessee, Knoxville; and [3]Lawrence Berkeley National Laboratory, Berkeley, Calif.

http://enigma.lbl.gov

Project Goals: Bioremediation strategies involving *in-situ* microbial stimulation for Cr(VI) -reduction/immobilization are ongoing, but determining their relative success is complex. By using controlled laboratory conditions, the influence of individual variables on the successful community structure, dynamics and the metal-reducing potential can be studied. The goal of the current work was to determine the impact of lactate stimulation during sulfate limitation on the succession of a native microbial community. Triplicate anaerobic, continuous-flow reactors were inoculated with Cr(VI) contaminated groundwater from the Hanford, Washington 100-H area and incubated for 95 days to obtain stable, enriched communities. The microbial community structure shifted with a significant loss of diversity. Final communities were dominated by *Pelosinus* spp. and to a lesser degree, *Acetobacterium* spp. with small levels of other organisms including methanogens. The resultant diversity decreased from 63 genera within 12 phyla to 11 bacterial genera (from 3 phyla) and 2 archaeal genera (from 1 phylum). Isolation efforts attained four new strains of *Pelosinus* spp. Three of the 4 *Pelosinus* strains were capable of Cr(VI)-reduction and one also reduced U(VI). Under the tested conditions of limited sulfate, it appears that the sulfate-reducers, including *Desulfovibrio* spp., were outcompeted even though they are capable of fermentative growth. These results suggest that lesser-known organisms, such as *Pelosinus* spp., may play a more important role in metal-reduction than currently suspected.

Currently, a similar bioreactor experiment is underway and builds from the above findings. Hanford groundwater was enriched in duplicate bioreactors with lactate and; 1) no Cr(VI) to emulate the earlier work and establish a baseline, 2) 0.1 mg/L Cr(VI) to reflect the low plume concentrations nearing the Columbia river and 3) 3.0 mg/L Cr(VI) to reflect the source metal concentration. Multiple levels of analysis are now underway after a 105 day experiment. These include temporal measurements of; 16S rRNA gene pyrosequencing, gas and organic acid quantification by GC and HPLC, Geochip, Phylochip, metagenomics, metatranscriptomics, metaproteomics, soluble Fe(III), HFO, Cr(VI) and U(VI) reduction assays, metal uptake characteristics for 36 metals, small metabolite quantification, and fluorescent antibody cell counts for species of *Desulfovibrio*, *Geobacter*, *Methanococcus*, and *Pelosinus*. These results will help to determine if and how the added complexity of Cr(IV) will influence the microbial community structure and metabolism as well as revealing the any acclimation of the community to Cr and quantifying the relative reduction potential.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231. Oak Ridge National Laboratory is managed by University of Tennessee UT-Battelle LLC for the Department of Energy under Contract No. DE-AC05-00OR22725.

207

ENIGMA Environmental Microbiology: Single-Cell Analysis Platforms for Genomic Analysis of Uncultivable Environmental Microbes

Suzan Yilmaz[1]* (syilmaz@sandia.gov), Peng Liu,[1] Robert J. Meagher,[1] Yooli Light,[1] Adam P. Arkin,[2] Terry C. Hazen,[3] **Anup K. Singh**,[1] and Paul D. Adams[2]

[1]Sandia National Laboratories, Livermore, Calif.; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; and [3]University of Tennessee, Knoxville

http://enigma.lbl.gov

Project Goals: We are developing a pipeline for single cell genomics that utilizes FISH (fluorescence *in situ* hybridization) for targeting species of interest, FACS (fluorescence activated cell sorting) for high throughput isolation of single cells, and MDA (multiple displacement amplification) for production of sufficient DNA from single cells for genome sequencing. This pipeline is being used for a number of collaborative projects in ENIGMA.

Current metagenomic techniques (e.g., microarray or 16S rRNA sequencing) relying on pooled nucleic acids from lysed bacteria can independently measure metabolic activity and the species present, but cannot link the activity deterministically to species. We are developing high-throughput tools for studying bacteria one cell at a time, allowing us to unravel the complex dynamics of population, gene expression, and metabolic function in mixed microbial communities. Our approach includes FISH-based identification of desired species, enrichment by cell sorting, followed by single-cell encapsulation, whole genome amplification and sequencing. Encapsulation of bacteria in nanoliter plugs in particular allows us to scale down conventional (microliter-volume) assays, such as WGA, into much smaller reaction volumes better suited to the size of an individual microbe. We are using this pipeline to analyze water samples from DOE bioremediation sites (e.g., Hanford) to identify

keystone organisms and link their functions to species. Furthermore, we are also using our single-cell genomics pipeline to complement the metagenomic sequencing efforts in ENIGMA. Metagenomic sequencing typically fails to achieve complete assembly and metabolic reconstruction of individual genomes in a complex community. Single-cell sequencing, together with metagenomics, makes it possible to assemble genomes of novel uncultivated organisms.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

208

ENIGMA Environmental Microbiology: Metagenomics-Enabled Understanding of Microbial Communities at DOE Contaminated Sites

Jizhong Zhou[1,2*] (jzhou@ou.edu), Zhili He,[1] Ye Deng,[1] Aifen Zhou,[1] Ping Zhang,[1] Qichao Tu,[1] Joy Van Nostrand,[1] Hao Yu,[1] Zhou Shi,[1] James Voordeckers,[1] Yongjin Lee,[1] Rong Song,[1] Liyou Wu,[1] **Matthew W. Fields**,[3] **David A. Stahl**,[4] **Judy D. Wall**,[5] **Terry C. Hazen**,[6,7] **Adam P. Arkin**,[2] and **Paul D. Adams**[2] (PDAdams@lbl.gov)

[1]University of Oklahoma, Norman; [2]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [3]Montana State University, Bozeman; [4]University of Washington, Seattle; [5]University of Missouri, Columbia; [6]University of Tennessee, Knoxville; and [7]Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://enigma.lbl.gov

Project Goals: Although high throughput sequencing and associated (meta)genomic technologies provide an avenue to determine genetic and organismal diversity of an ecosystem, linking the genetic/population diversity to phenotypic diversity across different organizational levels (e.g., molecular, cellular, populations, communities and ecosystems) is extremely challenging. As a part of the overall ENIGMA goal, the ultimate aim of this project is to utilize (meta)genomic technologies to better understand the mechanistic connections between molecular-level interactions/processes and community-level processes/functions. The following three specific objectives have been pursued: (i) To determine adaptation and molecular mechanisms of *Desulfovibrio vulgaris* Hildenborough (DvH) in response to multiple environmental stresses; (ii) To understand microbial community functional diversity at U/Cr-contaminated sites and develop high-throughput functional gene arrays (FGAs) for microbial community analysis; and (iii) To determine the responses, interactions, mechanisms and dynamics of groundwater/sediment microbial communities to U/Cr contamination and bioremediation treatments.

Long-term experimental evolution of DvH. To better understand the mechanistic connections between molecular-level functions and community-level processes, experimental evolution has been carried out to determine molecular mechanisms of DvH in response to high salinity. Significantly increased salt resistance was observed in evolved DvH (eDvH) with increased biomass, higher growth rate and shorter lag phases. Whole genome sequencing of eDvH at 1200 generations revealed specific point mutations and deletions. Their contribution to increased salt resistance has been proven by mutagenesis and phenotype analyses. Glu and Ala significantly increased in eDvH. After 5000 generations, the final biomass and growth rate of eDvH in the medium with high salinity was similar to that in the medium without extra salt. To further investigate the dynamics of evolution, repeatability of the evolution, whole genome sequencing, fitness assay and site-directed mutagenesis are in progress.

GeoChip-based metagenomic technology development. We have developed the GeoChip 4.0 series (4.0-4.2) for characterizing microbial communities. The GeoChip 4.0 series are manufactured based on the NimbleGen microarray format. For example, GeoChip 4.0 contains 120,054 distinct probes, and covers 200,393 coding sequences for 539 gene families in different microbial functional processes. The StressChip subset contains 22,855 probes covering 79,628 gene sequences for 46 genes involved in microbial responses to environmental stresses (e.g., temperature, osmolarity, oxidative status, nutrient limitation). The specificity, sensitivity and quantification of the developed GeoChip 4.0 series were evaluated computationally and experimentally. High specificity was observed for both synthesized oligonucleotides and genomic DNA from pure strains; the sensitivity was estimated to be 0.5 μg of DNA; the log(signal intensity) vs. log(DNA concentration) was highly correlated (R = 0.925). All the results showed that the GeoChip 4.0 series are specific, sensitive, and quantitative tools for characterizing microbial communities.

GeoChip applications. GeoChips have been used to study groundwater microbial communities to examine sustained reduction of contaminants using slow-degrading/slow-hydrolysis e-donors. At the Oak Ridge site, a one-time injection of emulsified vegetable oil (EVO) was used to examine U(VI) bioreduction and immobilization. Samples collected from the control and treatment wells (W1-7) were analyzed using GeoChip 3.0. Acetate, from EVO biodegradation, stimulated $NO3^-$, Mn(IV), Fe(III), $SO4^{2-}$, and U(VI) bioreduction in W1-7 and increased functional gene diversity. After EVO depletion, functional gene diversity declined. Fe(III)- and sulfate-reducing bacteria could play key roles in U(VI) reduction, whereas acetogens, denitrifiers and methanotrophs could be important for e-donor production and maintaining favorable reducing conditions. At the Hanford site, a one-time injection of poly-lactate was used to test Cr(VI) bioreduction. The groundwater microbial communities were monitored for 390 days using GeoChip 4.0. Cr(VI) was effectively reduced and functional gene diversity increased. Fe(III)- and sulfate-reducing bacteria could play key roles in Cr(VI) reduction, whereas denitrifiers could be important for maintaining reducing conditions.

Metagenomic sequencing. We have sequenced or resequenced metagenomes and dominant isolates from Oak Ridge FRC wells FW106 (contaminated with uranium, nitric acid, organics, and mercury) and FW301 (pristine). The FW106 metagenome was previously sequenced, and the results suggest extensive lateral transfer of metal resistance and organic compound metabolism genes. To extend these analyses, the metagenomes of FW106 and FW301, and genomes of multiple isolates of the dominant *Rhodanobacter* strain found in FW106 were sequenced or resequenced using Illumina sequencing technology. We are currently conducting comparative analyses of FW106 to FW301 to identify ecological trends observed between pristine and highly stressed groundwater communities. Furthermore, we are comparing the FW106 metagenome to *Rhodanobacter* isolates to confirm predicted lateral transfer events.

Molecular ecological network analysis. A novel random matrix theory-based approach has been developed to construct molecular ecological networks (MENs) based on GeoChip hybridization or high-throughput sequence data. Various mathematic and statistical tools and methods have been integrated into a comprehensive MEN analysis pipeline (MENAP). We have applied this approach to construct and analyze MENs from the Oak Ridge EVO experiment described above. Functional MENs were constructed from three GeoChip datasets: (i) Early EVO injection (≤31 days), (ii) Days 80 to 140, and (iii) Control well (7 time points) and pre- injection (0 day). All three constructed functional MENs posed general network characteristics (scale-free, small world and modularity), and the topology of these functional MENs was distinctly different, suggesting that the interactions among different microbial functional groups/populations in each community were dynamically altered during uranium bioremediation. Additionally, the changes in network structure were significantly correlated with environmental geochemical dynamics and EVO concentrations.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

209

ENIGMA Environmental Microbiology: Effect of Nitrate Stress on Metal-Reducing Microbes and Results of Nitrate Push/Pull Field Tests at Hanford 100H

Romy Chakraborty[1]* (rchakraborty@lbl.gov), Sharon E. Borglin,[1] Boris Faybishenko,[1] Paramvir Dehal,[1] **Adam P. Arkin**,[1] **Terry C. Hazen**,[2,3] **Matthew W. Fields**,[4] Jil Geller,[1] Julian Fortney,[2,3] Dominique Joyner,[2,3] Mark Conrad,[1] and **Paul D. Adams**[1] (PDAdams@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [2]University of Tennessee, Knoxville; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]Montana State University, Bozeman

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project seeks to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. One goal is to understand the effect of environmental stressors that enable ENIGMA-relevant microorganisms to thrive in such environments.

As part of the ongoing investigation of sustainable bioremediation of Cr(VI) in groundwater at the Hanford 100H area, we performed groundwater biostimulation tests by injecting Hydrogen Release Compound (HRC) and three lactate (17mM) injection experiments. To investigate the response of resident microbes to nitrate stress, a push-pull test was then conducted by injecting 55 gals of groundwater (collected from the background well) with KNO_3 (nitrate concentration 5,000 ppm) in October, 2010. After one day, pumping began from the same well, and lasted for 16 days. As a result of nitrate injection, total biomass decreased and sulfate concentration increased, but the sulfide and iron concentrations dropped. During pumping, the nitrate concentration decreased about 3 orders of magnitude. PLFA data showed biomass on the order of 10^7 cells/ml prior to push pull, and dropping off to 10^5 cells/ml during the test, but recovering toward the end back to 10^7 cells/ml by the end. Biomarker lipids indicate a shift toward monoenoics indicating an increase in gram negative bacteria and decrease in branched lipids (gram postive) and branched monoenoic (sulfate reducers).

We discuss the field test results in details and elucidate the effect of nitrate stress on environmentally relevant microbes *Geobacter metallireducens* and *Desulfovibrio vulgaris* as observed through controlled lab experiments. The lab studies and the field study together will help in understanding the overall fate of microbes under changing environmental conditions in the field and the key cellular mechanisms impacted by such stress conditions.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

* Presenting author

210

ENIGMA Microbiology Physiology: Dissecting the Physiology and Community Interactions of Environmental Bacteria

Adam P. Arkin[1]* (APArkin@lbl.gov), Kelly Bender,[2] David Stahl,[3] Aindrila Mukhopadhyay,[1] Jan Liphardt,[1] Swapnil Chhabra,[1] Adam Deutschbauer,[1] Michael W.W. Adams,[4] Steve Yannone,[1] Mark Biggin,[1] Paramvir Dehal,[1] Marcin Joachimiak,[1] Kimberly Keller,[8] Birte Meyer,[3] James Moberly,[6] Morgan Price,[1] Lara Rajeev,[1] Trent Northen,[1] Matthew W. Fields,[5] Dwayne Elias,[6] Nitin Baliga,[7] **Judy Wall**,[8] and Paul Adams[1]

[1]Lawrence Berkeley National Laboratory; [2]Southern Illinois University; [3]University of Washington; [4]University of Georgia; [5]Montana State University; [6]Oak Ridge National Laboratory; [7]Institute for Systems Biology; and [8]University of Missouri

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project aims to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. The ENIGMA Laboratory Component focuses its efforts on uncovering the genetic and biochemical basis of physiology and metabolism in key microbial species in isolation and in laboratory-defined communities. These data are used to develop models of microbial community activity and principles of community organization in an effort to predict the role that microbial species and their interactions play in the dynamics of geochemical transformations in a changing environment.

The mission of the Laboratory Component involves four primary goals that are closely coordinated with the Environmental, Biotechnology, and Computational Components within ENIGMA. Here, we describe these goals and highlight recent progress in these areas with a focus on the biology and interactions of the sulfate-reducing bacterium (SRB) *Desulfovibrio vulgaris* Hildenborough (DvH) in microbial consortia (1). In collaboration with the Environmental Component, we will characterize key microbes and communities enriched directly from metal-contaminated sites and prioritize these systems for deep functional annotation based on physiological properties, interactions among microorganisms, and potential biogeochemical activities. With flexible experimental techniques developed in concert with the Biotechnology component, we will use metabolomics, proteomics, transcriptomics, high-resolution microscopy, and high-throughput mutagenesis/phenotyping approaches to define transcript architecture, signal transduction, gene function, sRNA function, protein interactions and localization, elucidate the structure and function of biofilm formation, the metalloproteome, metabolism and mechanisms of specific interaction in these key isolates and their relatives (2). With the Computational group, we will use these data to construct predictive models of metabolism and gene regulation, which in turn will be used as drivers for new hypotheses and further experimentation (3). Using key microbial isolates, we will constitute laboratory simulation of enriched and defined assemblages in planktonic and attached states and assess different cellular compositions under varying environmental constraints. As a two species community model system, we have used long-term evolution experiments to map the genetic basis of improved syntrophy between DvH and a methanogen. More recently, transcriptomics and whole-genome mutant fitness profiling have identified key genes and pathways involved in the syntrophy of *Desulfovibrio alaskensis* G20 with a methanogen (4). These form a basis for pilot studies on the molecular dissection of competitive association among multiple SRB and methanogens in common culture. Lastly, we aim to determine how information flows among microbes and to assess the specificity, mutualism, and antagonism among pairs and groups of microbes by application of genetic, biochemical and computational technologies.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

211

ENIGMA Microbial Physiology: Systematic Mapping of Two Component Response Regulators to Gene Targets in a Model Sulfate-Reducing Bacterium

L. Rajeev,[1] E. Luning, [1] K. Keller,[2] G. Zane,[2] P. Dehal,[1] M. Price, [1] A. Arkin, [1] J. Wall,[2] and **A. Mukhopadhyay**[1]* (amukhopadhyay@lbl.gov)

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]Biochemistry Division, University of Missouri, Columbia

http://enigma.lbl.gov/

Project Goals: The goal of this project is to elucidate the mechanisms by which bacteria connect core metabolic functions to environmental cues and stresses. Such signaling remains poorly understood even in the best-studied model organisms. As part of near-term goals, we developed methods to understand signal transduction pathways and the corresponding regulatory networks in the sulfate-reducing bacterium, *Desulfovibrio vulgaris*. Specifically we focused on two component signal transduction systems. In ongoing research our goal is to understand the physiological relevance of the regulatory maps we have discovered. Long-term goals of this project are to extend our methods to evaluate multiple organisms that coexist in an ecological niche and deduce the connections between the environment and a microbial community that exists in it.

Desulfovibrio vulgaris is an environmentally relevant bacterium that serves as a model system for dissimilatory sulfate reduction. It is an important member of anaerobic syntrophic communities and is of interest for its metal reduction ability. The strain Hildenborough encodes a large number of two component regulatory systems, none of which are characterized. We sought to map the transcriptionally acting response regulators of these signal transduction systems to their gene targets. In order to accomplish this goal, we developed an in vitro DNA-affinity-purified-Chip method. We successfully determined 200 gene targets for 24 response regulators, which constitute the majority of this class of regulators in *D. vulgaris*. Our results enabled functional predictions and the identification of binding site motifs for several regulators (1). As expected, several simplex and complex regulatory modules were discovered. Of these an important regulatory network uncovered in our study is centered on the lactate utilization pathway, which appears to be under the control of multiple response regulators. The regulators include a lactate-responsive, a nitrite-responsive, a phosphate-responsive regulator and a potential oxidative stress responsive regulator. Here we present the comprehensive set of regulatory maps obtained using the DAP-chip method. Further, we describe the results from our experiments to evaluate the response of *D. vulgaris* carbon utilization pathway to various stresses such as nitrite and phosphate.

Reference

1. Rajeev L, Luning EG, Dehal PS, Price MN, Arkin AP, Mukhopadhyay A* Systematic mapping of two component response regulators to gene targets in a model sulfate reducing bacterium. *Genome Biology* 2011

This work conducted by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

212

ENIGMA Microbial Physiology: Evidence-Based Annotation of Gene Function in Metal-Reducing and Sulfate-Reducing Bacteria Using Genome-Wide Mutant Fitness Data

Adam Deutschbauer[1]* (AMDeutschbauer@lbl.gov), Grant M. Zane,[2] Samuel Fels,[2] Morgan Price,[1] Jennifer V. Kuehl,[1] Kelly Wetmore,[1] Swapnil R. Chhabra,[1] Thomas R. Juba,[2] Gareth P. Butland,[1] **Adam P. Arkin**,[1] **Judy D. Wall**,[2] and Paul Adams[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif. and [2]University of Missouri, Columbia

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project seeks to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. One goal is to develop and apply high-throughput genetic technologies to identify causative genes and pathways rapidly in ENIGMA-relevant microorganisms.

The development of genetic tools to study the functions of thousands of poorly characterized genes in newly isolated environmental microbes is a rate-limiting step in determining the activity of the keystone organisms in communities. One promising approach to meet this challenge is the large-scale generation of mutant libraries and gene fitness assays as a complement to standard whole genome sequencing and classical annotations based on homology and computation. Currently, large-scale mutant collections have been assembled for only a small sample of typically pathogenic bacteria. With the availability of a universal, sequence-verified TagModule collection (Oh et al. NAR 38:e146, 2010) that has been introduced into a hyperactive mini-Tn*5* (Larsen et al., Arch Microbiol. 178:193, 2002), the creation of such libraries has been facilitated for a number of environmental microbes. We have generated sequenced and archived transposon mutant collections of 24,688 mutants of *Shewanella oneidensis* MR-1, 15,477 of *Desulfovibrio alaskensis* G20, and 13,007 of *Desulfovibrio vulgaris* Hildenborough. Simple scanning of the genes not mutated provides a limited list that includes essential genes for the condition of mutant isolation. Pools of tagged or barcoded transposon mutants of most non-essential genes can be used to probe the gene fitness in different growth conditions and when subjected to various stresses. Recent work with *S. oneidensis* (Deutschbauer et al., PLOS Genetics, 7:e1002385, 2011) has shown that many hypothetical genes and potentially redundant genes have detectable phenotypes when this approach is applied across a large number of diverse growth conditions. In some instances, specific molecular functions can be predicted (evidence-based annotations) where nearly meaningless annotations were the extent of previous knowledge.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

213

ENIGMA Laboratory: Evolutionary and Ecological Origins of Community Assembly, Stability, and Efficiency

Birte Meyer[1]* (birtemey@u.washington.edu), Kristina L. Hillesland,[1] Jason Flowers,[1] Nicolas Pinel,[2] Nicholas Elliott,[1] **Judy Wall**,[3] Marcin Joachimiak,[4] Aifen Zhou,[5] **Joe Zhou**,[5] Jennifer Kuehl,[4] **Adam Deutschbauer**,[4] Paramvir Dehal,[4] Morgan Price,[4] Zhili He,[5] **Aindrila Mukhopadhyay**,[4] Lara Rajeev,[4] Edward Baidoo,[4] **Trent Northen**,[4] **Nitin Baliga**,[2] **Mark Biggin**,[4] Ming Dong,[4] **Jap Bing**,[4] Peter Walian,[4] **Anup Singh**,[6] Suzan Yilmaz,[6]

Dwayne Elias,[7] **Matthew Fields**,[8] Hector Garcia-Martin,[4] **Terry Hazen**,[4] **Adam Arkin**,[4] and **David Stahl**[1]

[1]University of Washington, Seattle/Bothell; [2]Institute for Systems Biology, Seattle, Wash.; [3]University of Missouri, Columbia; [4]Lawrence Berkeley National Laboratory, Berkeley, Calif.; [5]University of Oklahoma, Norman; [6]Sandia National Laboratories; [7]Oak Ridge National Laboratory; and [8]Montana State University, Bozeman

http://enigma.lbl.gov

Project Goals: A goal of DOE and ENIGMA is to understand and ultimately predict microbialcommunity assembly and the adaptive response of communities to environmental change. To address these questions, we are examining assemblies of simple two-member communitiescomposed of a secondary fermenter like *Desulfovibrio* species and hydrogenotrophic methanogens. This type of association (syntrophy) is representative of a trophic interaction sustaining both pollutant transformation and organic matter mineralization in many anoxicenvironments typical of the subsurface. To study the metabolic versatility and specificity ofthese assemblies, we first investigated the association of twelve different pairings of *Dv.* species and two methanogenic species (*Methanococcus maripaludis* and *Methanospirillum hungatei*).The results demonstrated that different *Desulfovibrio*-methanogen pairings vary significantly in their growth characteristics, most notably in their ability to ferment lactate at elevated hydrogen levels, presumably reflecting differences in their syntrophy-associated enzyme systems (e.g., hydrogenases and electron transfer complexes). Those studies now serve to direct a systems-level approach to the study of common and divergent features of community interaction: focusing on the genetic and metabolic signatures of efficient species interaction, major determinants of community stability, and the capacity for these communities to improve through adaptive evolution.

Comparative studies have so far shown that both the electron transfer system and themediator for electron transfer differ among *Desulfovibrio* species. For example, comparison ofdifferent assemblies grown in chemostats under various dilution rates demonstrated theimportance of formate as a major mediator of electron exchange in *Dv. alaskensis* strain G20 syntrophic cocultures in contrast to the hydrogen exchange-based system of *Dv. vulgaris* strain Hildenborough. Notably, the transcript analyses revealed that gene expression during syntrophic growth of *Dv. alaskensis* str. G20 also varies with both growth rate and the methanogenic partner. These conclusions were subsequently confirmed using a tagged-transposon *Dv. alaskenis* G20 mutant mini-library (1200 strains) to examine the relative fitness of differentinsertion mutants grown syntrophically in chemostats. Complementary studies are examiningthe adaptive evolutionary response of the two species to syntrophic growth. Ongoing laboratory evolution experiments of 24 replicated lines have so far documented a remarkable capacity for rapid improvement in the stability and efficiency of this mutualism after only 1000 generationsof cooperative growth. The genetic basis of improved cooperation is now being examined by genome resequencing, initially of twelve of the pairings at 1000 generations using both Illuminaand SOLiD next generation sequencing platforms and microarrays. Since some lines have evolved to obligate syntrophy, the history of their adaptive evolution will be reconstructed using single cell genome sequencing of earlier generations. The first mutations in *Dv. vulgaris* to become fixed in multiple lines were in an outer membrane porin (DVU0799), suggesting thatamino acid replacements near the outer face of this porin alters the flux of metabolites and/or nutrients. Other notable and frequent mutations were in genes implicated in EPS synthesis and regulation of lactate metabolism. Collaborative biophysical studies with ENIGMA members arenow exploring altered function of the different porin mutants and mutants implicated in EPSsynthesis. Of particular note was the replacement in all evolved lines of six amino acids in thesensory PAS domain of a histidine kinase (DVU3022) implicated in the regulation of lactatemetabolism. The high frequency replacement with an identical short amino acid sequencehaving predicted sensory function suggests a novel phase variation-like mechanism of adaptive response. Together these ongoing studies point to both common and divergent mechanisms ofinterspecies interaction, and offer a framework to better resolve genetic, structural, and metabolicfeatures contributing to stability and efficiency of community assembly.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biologicaland Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC0205CH11231

214

ENIGMA Microbial Physiology: Cooperation Impacts Structure in a Syntrophic Biofilm of *Methanococcus maripaludis* and *Desulfovibrio vulgaris*

Kristen A. Brileya,[1,2] Matthew W. Fields[1,2]* (matthew.fields@erc.montana.edu), Adam P. Arkin,[3] and **Paul D. Adams**[3] (PDAdams@lbl.gov)

[1]Center for Biofilm Engineering, [2]Department of Microbiology, Montana State University; and [3]Lawrence Berkeley National Laboratory

http://enigma.lbl.gov

Project Goals: Elucidate the relationship between structure and function in a methanogenic biofilm that consists of a sulfate-reducing bacterium and methanogenic archaeon. The functions of interest are carbon-compound oxidation and methanogenesis between syntrophic populations. The structure refers to population organization within a self-assembled, biofilm community.

Transfer of reduced carbon and electrons between microbial community members is of interest in anoxic systems, and methanogenesis represents a crucial trophic level that can include sulfate-reducing bacteria and methanogenic

archaea. The current work uses a dual-culture approach to examine the structure of a syntrophic biofilm formed by the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough and the methanogenic archaeon *Methanococcus maripaludis.* Biofilm was grown in a continuously stirred reactor where cells could attach to a silica surface or remain suspended. Under the tested conditions, *D. vulgaris* formed monoculture biofilm, but *M. maripaludis* did not. However, *M. maripaludis* did form pellicles in static, batch cultures while *D. vulgaris* did not form a pellicle. Under syntrophic conditions, a methanogenic biofilm formed and reached steady-state in approximately 7 days based upon protein levels and methane mass flux. Biofilm establishment was dependent upon initial colonization by *D. vulgaris* that was followed by recruitment of *M. maripaludis* into the biofilm matrix. The initial *Desulfovibrio*:*Methanococcus* biofilm ratio was approximately 375:1 but steady-state biofilm reached a ratio of 4:1. Steady-state biofilm was fixed for Fluorescence *in situ* Hybridization (FISH) and confocal laser scanning microscopy (CLSM). FISH revealed a framework of *D. vulgaris* with both single cells and large micro-colonies of *M. maripaludis* interspersed throughout the biofilm. 3D-FISH and CLSM of hydrated intact biofilm confirmed steady-state biofilm irregularity, with ridge, valley and spire macro-architecture. Key structural signatures were observed that confirmed the cooperative nature of the community using a newly developed model. Colorimetric assays indicated cell-associated carbohydrate was composed of .035 μg hexose/μg protein, .017 μg pentose/μg protein and .011 μg uronic acid/μg protein, similar to *D. vulgaris* mono-culture biofilm and approximately 5 times less than *M. maripaludis* pellicles. Filaments presumed to be protein have been observed in dual-culture biofilm matrix with electron and atomic force microscopy, and matrix was sensitive to proteinase K treatment during preliminary work with Catalyzed Reporter Deposition FISH. Syntrophic biofilm 3-D structure appears to be initialized by *D. vulgaris* that provides an advantageous environment for *M. maripaludis* to establish micro-colonies throughout the *D. vulgaris* scaffold. The coculture biofilm growth mode resulted in a 10-fold higher methane production per *M. maripaludis* biomass than the planktonic only growth mode, and these results indicated that the structure of cooperative interactions between a bacterium and archaeon positively impacted function.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

215

ENIGMA Microbial Physiology: Assimilatory and Dissimilatory Metallomics of *Desulfovibrio vulgaris* and *Pelosinus* Strain A11

W. Andrew Lancaster* (lancaste@uga.edu),[1] Israel Scott,[1] Brian Vacarro,[1] Angeli Lal Menon,[1] Farris L. Poole,[1] Jil Geller,[2] Jennifer J. Mosher,[3] Terry C. Hazen,[2,3] Dwayne A. Elias,[3] Michael W.W. Adams,[1] Adam P. Arkin[2] and Paul Adams[2]

[1]University of Georgia, Athens; [2]Lawrence Berkeley National Laboratory / ENIGMA, Berkeley, Calif.; and [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.

http://enigma.lbl.gov/

Project Goals: The goals of the ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project are to understand at a molecular systems biology level the microbial communities at DOE sites contaminated with heavy metals or radionuclides with sufficient detail to predictively model interactions within microbial and community processes that drive complex geochemistry in key environments. We expect to define biological principles governing selection of microbial community function and composition in given environments.

While some heavy metals are environmental contaminants, metals in general have a very positive role in biological systems as they afford proteins virtually unlimited catalytic potential, enable electron transfer reactions and greatly impact protein stability. Consequently, metal-containing proteins play key roles in virtually all biological processes. However, the full complement of metalloproteins within a given cell cannot be predicted solely from bioinformatic analyses of a genome sequence since metal coordination sites are diverse and poorly recognized. Hence it is not possible to predict the number and types of metals that a microorganism utilizes and how these might vary with the growth conditions. Determining the metals that are taken up during microbial growth and the metal content of fractionated native biomass can provide insight into these issues and, using coupled MS/MS analyses with extensive fractionations, can reveal completely new aspects of metal metabolism (1). Herein we report comparative metallomic analyses of the model microbe, *Desulfovibrio vulgaris* strain Hildenborough (DvH), and a newly described microbe, *Pelosinus* strain A11, which was recently isolated from an enrichment of a groundwater sample from the Hanford site by Elias and coworkers at ORNL. DvH was grown on lactate under sulfate-reducing conditions in a 600-liter metal fermenter and in a glass 5-liter fermenter and the nature of metals assimilated were compared. *Pelosinus* A11 was also grown at the 600-liter scale on fructose in the absence of sulfate and the assimilated metals were determined using ICP-MS (53 elements). Both organisms assimilated 20 or so metals into their cytoplasmic fractions, but the types and amounts were species specific. For example, DvH assimilated cobalt, cadmium and tungsten into high molecular weight complexes (>3 kDa) but *Pelosinus* A11 did not. In contrast to DvH, *Pelosinus* A11 assimilated vanadium, copper and uranium into >3kDa complexes. A number of additional factors affecting metal metabolism were analyzed, including growth with and without added chromium or uranium to study assimilation and dissimilatory reduction of these metals. The results will be presented in terms of the ranges and types of metals assimilated by DvH and *Pelosinus*

A11 and the metals available in the organisms' natural environments, including Hanford groundwater.

Reference

1. Cvetkovic, A., Menon, A. L., Thorgersen, M., Scott, J. W., Poole, F. L., Jenney, F. E., Lancaster, W. A., Praismann, J. L., Shanmukh, S., Vaccaro, B., Trauger, S. A., Kalisiak, E., Apon, J. V., Siuzdak, G., Yannone, S. M., Tainer, J. A. and Adams, M. W. W. (2010) Microbial metalloproteomes are largely uncharacterized. Nature 466, 779-782

This work was conducted as part of the ENIGMA project and was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

216

ENIGMA Microbiology Physiology: Accurate, High-Throughput Identification of Stable Protein Complexes in *Desulfovibrio vulgaris*

John-Marc Chandonia[1,3]* (JMChandonia@lbl.gov), Maxim Shatsky,[1,3] Ming Dong,[1] Haichuan Liu,[2] Lee Yang,[1] Jil T. Geller,[1] Megan Choi,[1] Barbara Gold,[1] Nancy L. Liu,[1] Marjon Khairy,[1] Sonia Reveco,[1] Tom R. Juba,[5] Bonita R. Lam,[1] Evelin D. Szakal,[2] Simon Allen,[2] Sunil Kumar,[4] Farris L. Poole,[4] **Steven E. Brenner**,[1,3] **Steven C. Hall**,[2] **Susan J. Fisher**,[1,2] **Michael Adams**,[4] **Terry C. Hazen**,[6] **Judy D. Wall**,[5] **Swapnil Chhabra**,[1] **Jian Jin**,[1] **H. Ewa Witkowska**,[2] **Adam P. Arkin**,[1,3] **Gareth P. Butland**,[1] **Mark D. Biggin**,[1] and **Paul D. Adams**[1]

[1]Lawrence Berkeley National Laboratory; [2]University of California, San Francisco; [3]University of California, Berkeley; [4]University of Georgia; [5]University of Missouri, Columbia; and [6]Oak Ridge National Laboratory

http://enigma.lbl.gov

Project Goals: *Desulfovibrio vulgaris* has been selected as a model bacterium for intensive study by ENIGMA because it can reduce heavy metals and radionuclide contaminants present in the soil at many DOE sites, rendering the contaminants insoluble. ENIGMA seeks to model, at a molecular systems level, how this and similar bacteria respond to natural and human induced changes in their environment and how this alters their ability to stabilize contaminants in the soil. A component of our strategy is to develop and use high throughput pipelines to purify and characterize soluble protein complexes. We expect that these interaction data will improve our ability to produce accurate metabolic and regulatory models of key members of microbial communities.

The group led by Mark Biggin has developed a novel method for identification of stable, soluble protein complexes in microbes. In a small-scale pilot study, we showed that many protein complexes survive intact through a series of orthogonal chromatographic methods, with complex components having correlated elution profiles. These profiles were measured with the aid of mass spectrometry (MS) and iTRAQ reagents (Dong et al., 2008). We developed statistical and machine learning methods to analyze a full-scale data set, which were required in order to obtain biologically meaningful results due to the high potential for false positives (FP) caused by co-elution of proteins that are not part of a complex. Our methods were tuned using a manually curated gold standard (GS) set. As a first high-throughput study, we demonstrated this technique in identifying a high-precision subset of stable protein complexes in *Desulfovibrio vulgaris*. We have shown that our predicted network of interactions is significantly enriched in pairs with similar functional annotations. The quantitative information from elution profiles allowed us to develop a statistical model to estimate the false discovery rate in our predictions; because this varies according to how "crowded" the eluted fractions are, we are able to identify a subset of hundreds of highly reliable (i.e., with very low false discovery rate) interactions, as well as a much larger set of interactions that can be predicted with known false discovery rates. Advantages of the tagless approach include not requiring a mutant library (needed for alternative tag-based approaches such as TAP), and a false discovery rate comparable to TAP.

The group led by Gareth Butland has identified a number of protein complexes using TAP. We have developed an automated pipeline for synthesis of tagged gene constructs in collaboration with Swapnil Chhabra (Chhabra et al., 2011). To date, over 700 pulldowns (comprising more than 600 unique *D. vulgaris* strains) have been subject to analysis. In these experiments, more than 10,000 interactions were detected with over 1,000 distinct prey proteins. Using curated GS datasets (as in the tagless analysis), we filtered out ubiquitous proteins and other likely FP, resulting in a set of high-confidence interactions. A number of these interactions have been reciprocally confirmed, using strains in which the original prey protein was tagged and used as bait. Preliminary analysis of the data have identified several novel complexes, including multiple paralogous versions of the DnaJK-GrpE chaperone complex, each of which is bound to a small protein that may act as an allosteric regulator.

References

1. Chhabra SR, Butland GP, Elias D, Chandonia JM, Fok OY, Juba T, Gorur A, Allen S, Leung CM, Keller K, Reveco SA, Zane GM, Semkiw ES, Prathapam R, Gold B, Singer M, Ouellet M, Szakal ED, Jorgens D, Price MN, Witkowska HE, Beller HR, Arkin AP, Hazen TC, Biggin MD, Auer M, Wall JD, Keasling JD. 2011. Generalized Schemes for High Throughput Manipulation of the *Desulfovibrio vulgaris Hildenborough* Genome. *Applied and Environmental Microbiology 77*:7595-7604.
2. Dong M, Yang LL, Williams K, Fisher SJ, Hall SC, Biggin MD, Jin J, Witkowska HE. 2008. A "tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J Proteome Res 7*:1836-1849.

ENIGMA is a Lawrence Berkeley National Laboratory Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Foundational Science through contract DE-AC02-05CH11231.

217

ENIGMA Microbiology Physiology: A Revised Bioenergetic Model of *Desulfovibrio vulgaris* Strain Hildenborough

Romy Chakraborty[1]* (RChakraborty@lbl.gov), Marcin Joachimiak,[1] Aifen Zhou,[2] Julian L. Fortney,[1] Zhili He,[2] Paramvir Dehal,[1] Morgan R. Price,[1] Jizhong Zhou,[2] Terry C. Hazen,[3] Adam P. Arkin,[1] Judy D. Wall,[4] Adam Deutschbauer,[1] **Swapnil R. Chhabra**,[1] and Paul Adams[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, Calif., [2]University of Oklahoma, Norman; [3]Oak Ridge National Laboratory, Oak Ridge, Tenn.; and [4]University of Missouri, Columbia

http://enigma.lbl.gov

Project Goals: The ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) project seeks to elucidate the mechanisms and key processes that enable microorganisms and their communities to function in metal-contaminated soil sites. One goal is to understand the bioenergetic processes that enable ENIGMA-relevant microorganisms to thrive in such environments.

Sulfate reducing bacteria (SRB) play important roles in global sulfur and carbon cycling through their ability to completely mineralize organic matter while respiring sulfate to hydrogen sulfide. They are ubiquitous in anaerobic environments and have the ability to reduce toxic metals such as Cr(VI) and U(VI). While SRB have been studied for over three decades, bioenergetics of this clade of microbes is poorly understood. *Desulfovibrio vulgaris* strain Hildenborough (DvH) has served as a model SRB for the last several years with the accumulation of transcriptomic, proteomic and metabolic data under a wide variety of stressors. To further investigate the three proposed modes of energy generation in this anaerobe we conducted a systematic study involving multiple electron donor and acceptor combinations for growth. DvH was grown at 37°C in a defined medium with (a) 60mM Lactate + 30mM Thiosulfate, (b) 60mM Lactate + 40mM Sulfite (c) 60 mM lactate, 30 mM sulfate, (d) 120mM Pyruvate + 30mM Sulfate, (e) H_2 + 10mM Acetate + 30mM Sulfate, (f) formate 100mM + 10mM acetate + 25mM sulfate. g) no lactate, 5 mM acetate, 30 mM sulfate. Cells were harvested for transcriptomics at mid-log phase of growth for all conditions, when the optical density at 600nm was in the range 0.42-0.5. Initial results indicate that in spite of different electron donors, cells grown with sulfate, thiosulfate or sulfite as the electron acceptor show commonalities in gene expression and cluster together. Cells grown with H_2 and formate as electron donors as well as by pyruvate fermentation form a separate cluster consistent with differential pathway utilization. It was also observed that the ability of DvH to reduce metals like Cr(VI) when grown with lactate was greater than with formate as electron donor.

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

218

Berkeley Synchrotron Infrared Structural Biology (BSISB) Program: A Bioanalytical Facility for Characterizing Chemical Changes in Living Cells

Hoi-Ying N. Holman[1]* (hyholman@lbl.gov), Michael C. Martin,[2] and Peter Nico[1]

[1]Ecology Department and [2]Advanced Light Source, Lawrence Berkeley National Laboratory

Project Goals: A DOE Structural Biology User Facility for accelerating biological research.

The Berkeley Synchrotron Infrared Structural Biology program (BSISB) is funded by the Department of Energy to provide support to researchers whose work can benefit from real-time measurements and imaging of chemical changes in living cells while biological processes are underway. The BSISB program has developed synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy, which is ideally suited to measure at a chemicallevel how microbes adapt to changing environments. SR-FTIR spectromicroscopy is similar to the well-established non-invasive molecular analytical technique Fourier transform infrared (FTIR), except SR-FTIR is capable of discovering molecular information with micrometer spatial accuracy and signal-to-noise ratios 100 to 1000 times greater. Aqueous environments hinder SR-FTIR's sensitivity to bacterial activity, but the recent development of integrated in situ open-channel microfluidic culturing platforms circumvents the water-absorption barrier. These platforms enable real-time chemical imaging of bacterial activity in biofilms and facilitate comprehensive understanding of structural and functional dynamics in a wide range of microbial systems. Recently, the BSISB program has added visible (VIS) hyperspectral/fluorescence microscopy to its multi-modal imaging capability. Users can simultaneously track changes in cellular morphology, structure, and biological processes like gene expression and signaling during SIR experiments. The presentation will include examples that demonstrate the use of BSISB in wide ranging research projects that are important to DOE missions.

219

Structural and Functional Properties of Transporter Solute Binding Proteins

Sarah Zerbs[1]* (szerbs@anl.gov), Lisa M. Miller,[3] Ruth Pietri,[3] Jamey C. Mack,[1,2] Andrew Binkowski,[1,2] Changsoo Chang,[1,2] Marianne Cuff,[1,2] Yao Fan,[1,2] Karolina Michalska,[1,2] Jureck Osipiuk,[1,2] Kemin Tan,[1,2] Andrzej Joachimiak,[1,2] and **Frank Collart**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, Ill.; [2]Midwest Center for Structural Genomics, Argonne National Laboratory, Lemont, Ill.; and [3]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, N.Y.

http://www.bio.anl.gov/molecular_and_systems_biology/proteins.html

Project Goals: The Argonne "Environment Sensing and Response" Scientific Focus Area (ESR-SFA) program addresses the hypothesis that cellular behavior can be modeled through an understanding of the biological interface with the environment and the cellular responses that originate from the cell/environment interaction. The overall objective of the ESR-SFA program is to identify the molecular basis of cellular transport and sensory pathways that mediate the response to environmental nutrients. We recently identified four possible solute binding proteins (SBPs) in *R. palustris* that demonstrated binding to aromatic lignin monomers. We used molecular crystallography to examine the molecular basis for the aromatic compound binding specificity of these proteins and assess the feasibility for the development of sequenced based prediction methods to determine ligand binding specificity.

Rhodopseudomonas palustris metabolizes aromatic compounds derived from lignin-degradation products and has the potential for bioremediation of xenobiotic compounds. In this study, we exploit the ligand binding specificity of solute binding proteins (SBPs) to evaluate the transport specificity of ABC transporters for lignin degradation products for multiple transporters. We recently identified four possible SBPs in *R. palustris* CGA009 that demonstrated binding to aromatic ligands similar to lignin monomers. To expand on these initial results, we selected homologues from *R. palustris* strains HaA2 and BisB5, *B. japonicum* USDA110 and *S. meliloti* 1021 and subjected targets to a directed screen against potential aromatic ligands. A fluorescent thermal shift (FTS) assay was used to screen for ligands of purified SBPs using a library of 46 compounds. Individual proteins exhibited binding to multiple structurally similar molecules, but also demonstrated clear ligand preferences arising from carbon chain length and ring modifications. Arrangement of FTS data by reference SBP group and common chemical features of the ligands enables organization of the binding capabilities into distinct groups of protein-ligand interactions. Overlay of the chemical specificity on an unrooted phylogenetic tree suggests a consistent relationship of the binding profile inferred by the FTS assay with the phylogenetic relatedness.

To establish a molecular basis for the aromatic compound binding specificity of these proteins, 13 proteins were produced in a bacterial expression system, purified and routed to crystallization screens in the presence and absence of ligands. Each protein was screened with either no ligand (apo) or one of two preferred ligands (ligand1, ligand 2) indicated by induction of the highest melting temperature stabilization in FTS assays. These 39 combinations yielded multiple crystals, 9 of which gave refraction patterns suitable for solving structures. Solved structure resolutions are generally less than 2 Å and all structures indicate SBPs in a closed formation with a ligand in the binding pocket. Most targets were solved as monomers, however two targets exhibited dimers in the space group. In cases where an FTS-identified ligand was found in the active site, structure resolution was high enough to calculate hydrogen bond coordination in the residues that comprise the binding pocket. Based on these structures, ligand coordination differs between the two groups described by phylogenetic and *in vitro* functional analysis of these targets.

The crystal structures of these novel aromatic binding proteins from *R. palustris* confirmed the inferred functional assignment as binding proteins for aromatic compounds. The variances in binding pocket geometry and thermodynamic parameters for these SBPs are consistent with the observed sequence variance and sequence phylogeny for this protein family. This information will enable an understanding of the molecular mechanisms that underlie microbial interactions and thereby facilitate the development of system-level models predictive of cellular response to environmental conditions or changes.