

Systems Environmental Microbiology

36

The Virtual Institute of Microbial Stress and Survival VIMSS:ESPP Overview

Adam P. Arkin^{1,2,4,5*} (aparkin@lbl.gov), **Terry C. Hazen**^{1,2*}, Carl Abulencia^{1,7}, **Eric J. Alm**^{1,3}, Gary Anderson^{1,2}, Manfred Auer^{1,2}, Edward Baidoo^{1,2}, **Kelly S. Bender**^{1,12}, Peter Benke^{1,2}, Sharon Borglin^{1,2}, Eoin Brodie^{1,2}, Steven Brown^{1,6}, Lauren Camp^{1,2}, Romy Chakraborty^{1,2}, Swapnil Chhabra^{1,2}, Gabriela Chirica^{1,10}, Dylan Chivian^{1,2}, Michael Cipriano^{1,2}, Paramvir S. Dehal^{1,2}, Todd DeSantis^{1,2}, Elliot Drury^{1,8}, **Inna Dubchak**^{1,2}, Dwayne Elias^{1,8}, **Matthew W. Fields**^{1,10}, Veronica On Yi Fok^{1,2}, Julian Fortney^{1,2}, Sara Gaucher^{1,13}, Jil Geller^{1,2}, Masood Hadi^{1,13}, Zhili He^{1,11}, Chris Hemme^{1,13}, Kristina Hillesland^{1,9}, Hoi-Ying Holman^{1,2}, Katherine H. Huang^{1,2}, Y. Wayne Huang^{1,2}, Chiachi Hwang^{1,10}, Janet Jacobsen^{1,2}, Marcin P. Joachimiak^{1,2}, Dominique Joyner^{1,2}, **Jay Keasling**^{1,2,4}, Keith Keller^{1,2}, Kimberly L. Keller^{1,8}, **Martin Keller**^{1,6}, Yooli Light^{1,13}, Eric Luning^{1,2}, Robert Meagher^{1,13}, **Aindrila Mukhopadhyay**^{1,2}, Anthony Palumbo^{1,6}, Richard Phan^{1,2}, Tommy Phelps^{1,6}, Francesco Pingitore^{1,2}, Mircea Podar^{1,6}, Morgan N. Price^{1,2}, Alyssa Redding^{1,2}, Jarrod Robertson^{1,8}, Rajat Sapro^{1,13}, Christopher Schadt^{1,6}, Matthew Shirley^{1,8}, Amy Shutkin^{1,2}, Mary Singer^{1,2}, **Anup Singh**^{1,13}, **David A. Stahl**^{1,9}, Sergey Stolyar^{1,9}, Anitha Sundararajan^{1,10}, Yinjie Tang^{1,2}, Joy Van Nostrand^{1,11}, Sandra Villa^{1,2}, Christopher Walker^{1,9}, **Judy D. Wall**^{1,8}, Zamin Koo Yang^{1,6}, Huei-che Yen^{1,8}, Grant Zane^{1,8}, Aifen Zhou^{1,11} and **Jizhong Zhou**^{1,11}

¹Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov/>; ²Lawrence Berkeley National Laboratory, Berkeley, California; ³Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts; ⁴University of California, Berkeley, California; ⁵Department of Bioengineering, University of California, Berkeley, California; ⁶Oak Ridge National Laboratory, Oak Ridge, Tennessee; ⁷Diversa, Inc., San Diego, California; ⁸University of Missouri, Columbia, Missouri; ⁹University of Washington, Seattle, Washington; ¹⁰Montana State University, Bozeman,

—
GTL

Washington; ¹¹University of Oklahoma, Norman, Oklahoma; ¹²Southern Illinois University; and ¹³Sandia National Laboratories, Livermore, California

Project Goals (Abstracts 36–62): The Virtual Institute of Microbial Stress and Survival (<http://vimss.lbl.gov>) was established in 2002 through DOE Genomics:GTL funding of the Environmental Stress Pathway Project (ESPP). Using stress-response and metal reduction as target processes, and the environmentally important sulfate reducing bacteria (specifically *Desulfovibrio vulgaris* Hildenborough) as an organismal focus, ESPP is developing this system in two main thrusts: **1. Environmental characterization of biogeochemistry and microbial processes at metal/radionuclide contaminated sites, and 2. Laboratory and computational characterization of environmental stress pathways in microbes.** Phase 2 of this project is increasing focus on molecular determinants of community activity, stability and ecology. This research is managed by three core teams: Applied Environmental Microbiology Core, Functional Genomics and Imaging Core, and Computational and Systems Biology Core.

DOE oversees 350 cleanup projects involving soil contaminated with metals/radionuclides. The life-cycle cost of these projects is at least \$220 billion over 70 years, without breakthroughs. A thorough understanding of the biogeochemistry, especially stress responses in metal/radionuclide bacteria, enables prediction of natural attenuation and new strategies for remediation saving DOE billions in cleanup, risk assessment, and environmental stewardship. This application is representative of an array of environmental, ecological, and bioenergy stewardship challenges that rest on developing a detailed understanding of environmental microbial physiology, community interactions, population genetics and functions and ultimately evolution. The diversity of knowledge/technology necessary to accomplish these goals necessitates a team science approach, building a sophisticated experimental and computational infrastructure.

In the first phase of this project, we have succeeded in bringing a member of an important environmental class of microbe, the sulfate reducing bacteria (SRB), to nearly model organism status. In the last five years, we have been able to gain great insight into the stress response and metabolism of *Desulfovibrio vulgaris* Hildenborough (DvH). To do so we have developed an efficient genetic system; created a general purpose microbial systems biology pipeline; measured a wide range of genome-scale

* Presenting author

physiological responses to perturbations found in DOE contaminated environments; inferred regulatory systems and their evolution and conservation across metal-reducing SRB; and have just begun to probe DvH interactions with community members at DOE contaminated sites. The ESPP team has gelled as a mature working group of scientists able to balance the scale and coordination necessary for environmental systems biology with the special individual expertise to pursue detailed follow up projects. This synergy among labs and resources, and the economy of scale gained by building and utilizing a common experimental and computational infrastructure, enables attacking more complex and larger problems than any one or two labs could alone. This phase of the project has produced over 104 publications while team members have collectively made several hundred presentations of this work worldwide. The initial project has resulted in the development of a number of new technologies for functional genomics, data analysis, and physiological control of culture conditions. We have developed two community computational resources for comparative functional analysis of microbial genomes, MicrobesOnline (<http://microbesonline.org>) and RegTransBase (<http://regtransbase.lbl.gov>). We have 27 posters at this meeting outlining a number of the particular successes of this project as well as some of the work in preparation for phase two.

In phase two we will use this well-developed foundation to launch an expedition to discover the molecular mechanisms by which microbial community structure, function, and stability affects stress response and activity in the lab and in the field. With our current results in hand, we are poised to link field experiments strongly to controlled experiments with constructed consortia in the laboratory. We will continue to enrich our knowledge of the cellular networks of DvH and its relatives and expand our study to include its association with the methanogens with which it forms syntrophic associations. In stable co-cultures of *Desulfovibrio*-like organisms (Dvlo) and methanogens, we will discover the molecular basis of their metabolic coupling and other factors that lead to their more or less stable and active association leading to population growth, metal-reducing activity and resource utilization. We will construct more complex consortia to study the role of functional diversity, redundancy, and stress response in creating more or less stable microecologies via interactions at the community level. Finally, we will observe related microbes at DOE metal contaminated sites before, during, and after stimulation/stress to determine the differences and similarities to our laboratory experiments. There is little understanding of the molecular basis of microbial community formation, stability, resilience, specificity, activity and ecology, and

we will be carrying out one of the first large-scale systems biology effort to explore these ubiquitous and important phenomena. To do so we are also expanding our characterization tools to include high throughput genetic techniques, new metabolomics methods for mixed cultures, new profiling technologies for understanding population and gene expression changes (functional metagenomics) over time in complex field studies, as well as the computational tools with which to manage, analyze and interrelated these data sets.

This new effort to discover the molecular determinants of community activity, stability, and ecology (MDCASE) is truly only possible at this scale of collaboration with nine institutions and over 75 full time scientists working together across the scales from single microbe to /in situ /environmental community function. This presents its own challenges in project management, communication, and balancing individual interests against project wide studies. Over the last five years we believe we have established an agile style that allows for investigators to pursue individual research while benefiting from and contributing to group projects and culture. In depth posters on the three cores and their accomplishments as well as the project management experience can be found nearby.

37 GTL Applied Environmental Microbiology Core Research on Stress Response Pathways in Metal-Reducers VIMSS:ESPP

Terry C. Hazen^{1,2*} (TCHazen@lbl.gov), Carl Abulencia,^{1,4} Gary Anderson,^{1,2} Sharon Borglin,^{1,2} Eoin Brodie,^{1,2} Steve van Dien,⁸ Matthew Fields,^{1,7} Julian Fortney,^{1,2} Jil Geller,^{1,2} E. Hendrickson,⁵ Hoi-Ying Holman,^{1,2} J. Leigh,^{1,6} T. Lie,^{1,6} Richard Phan,^{1,2} Janet Jacobsen,^{1,2} Dominique Joyner,^{1,2} Romy Chakraborty,^{1,2} **Martin Keller**,^{1,3} **Aindrila Mukhopadhyay**,^{1,2} Christopher Schadt,^{1,3} **David Stahl**,^{1,6} Sergey Stolyar,^{1,6} Chris Walker,^{1,6} **Judy D. Wall**,^{1,5} Eleanor Wozel,^{1,2} Zamin Yang,^{1,3} Huei-che Yen,^{1,5} Grant Zane,^{1,5} and **Jizhong Zhou**^{1,9}

¹Virtual Institute of Microbial Stress and Survival (<http://vimss.lbl.gov>); ²Lawrence Berkeley National Laboratory, Berkeley, California; ³Oak Ridge National Laboratory, Oak Ridge, Tennessee; ⁴Diversa, Inc., San Diego, California; ⁵University of Missouri, Columbia, Missouri; ⁶University of Washington, Seattle,

* Presenting author

Washington; ⁷Montana State University, Bozeman, Montana; ⁸Genomatica, San Diego, California; and ⁹University of Oklahoma, Norman, Oklahoma

Project Goals: See goals for abstract 36.

Field Studies

Environmental Characterizations. Clonal libraries were used to characterize changes in community structure along a contaminant plume (Oak Ridge, TN) in terms of phylogenetic, functional, and geochemical changes. Such studies are essential to understand how a microbial ecosystem responds to perturbations. Relationships between community diversity and ecosystem function were idiosyncratic, and these results suggested the population distributions depended on conditions under which the local landscape was investigated. Principal component analysis showed that nitrate, uranium, sulfide, sulfate, and COD were strongly associated with particular bacterial populations. Sequences closely related to nitrate-reducing bacteria were predominant during the initial phase of the remediation process, but sequences representative of sulfate-reducers (*Desulfovibrio* and *Desulfosporosinus* spp.) and metal-reducers (*Geobacter* spp.) were detected at higher levels as uranium levels declined. When engineering controls were compared to the community structure and composition via canonical ordinations, population distributions could be related with dissolved oxygen control and the presence of bio-stimulant. During the biostimulation, population distributions followed geochemical parameters, and these results indicated that bacteria exhibited distributions at the landscape scale in concordance with predictable geochemical factors. The data indicated that relationships between community structure and ecosystem function were idiosyncratic, but temporal and spatial concordance were eventually observed for the two bio-stimulated wells. In addition, evolution of the communities suggested that many populations were initially present and the changing geochemical conditions selected for multiple populations in a non-clonal fashion. Ultimately, sequences associated with sulfate-reducing populations predominated, and enrichment cultures are dominated by *Desulfovibrio* populations.

Previous research specifically points toward SRB as environmentally relevant experimental systems for the study of heavy metal and radionuclide reduction, and our recent data has detected *Desulfovibrio* sequences at the FRC and Hanford 100H. To effectively immobilize heavy metals and radionuclides, it is important to understand the cellular responses to adverse factors observed at contaminated subsurface environments, such as mixed contaminants and the changing ratios of electron donors and acceptors. In a recent study, we focused on responses

to Cr(VI). At relatively low sulfide levels, *D. vulgaris* cells decoupled lactate oxidation from sulfate reduction for extended time periods even though all the Cr(VI) had been reduced. The cells could be protected by molecules that could complex Cr(III), and these results suggested that the reduction of Cr(VI) and/or reduction by-products were toxic to the cells. In addition, transcriptomic and physiological responses to Cr(VI) suggested that oxidizing stress was not a significant portion of the Cr induced stress, but was most likely a denatured protein response.

Technique Development for Environmental DNA and mRNA analysis. This year we did further optimization of the MDA approach to isolate and amplify DNA from samples with extremely low biomass. We used Hanford 100H samples to construct environmental libraries for sequencing and screening. We also evaluated three different methods in pure cultures to remove rRNA and tRNA from samples in order to screen mRNA expression that will eventually be applied to environmental samples. The first method utilizes biotin-modified oligos complementary to conserved regions in 16S and 23S rRNA and subtractive hybridization with streptavidin-coated magnetic beads. The second uses a commercially available exonuclease that specifically digests rRNAs bearing a 5' monophosphate group. The third method uses two rounds of reverse transcription, where rRNAs are first reverse transcribed with multiple universal primers for 16S and 23S RNAs, subsequently the RNA/DNA hybrids and cDNA are removed by sequential digestion with RNaseH and DNaseI, and the enriched mRNAs are then reverse transcribed using random primers. We evaluated these three methods by comparing disappearance of the 16S and 23S bands via electrophoresis, and their effect on mRNA quality and quantity by analysis of transcription levels of control (total RNA) vs. enriched mRNA as measured whole genome microarray. Enriched mRNAs from the first two methods generated more genes with altered transcript levels compared to untreated total RNA, with 19 genes (0.5%) for the exonuclease method and 74 genes (2%) for subtractive hybridization exhibiting significant differences ($P < 0.05$).

Genome Sequence. The genome sequence for *Desulfovibrio vulgaris* strain DePue has been closed and annotated in collaboration with other ESPP investigators. This genome appears highly conserved and syntenous compared with *D. vulgaris* Hildenborough except for dramatic differences in bacteriophage content. Additionally, strain DePue contains a large (~50 kb) unique region encoding for an exopolysaccharide production, modification and transport system. The genome of *Anaeromyxobacter* fw109 was recently completed and annotation

* Presenting author

is underway. *Anaeromyxobacter* fw109-5 is a mesophilic, iron-reducing δ -*Proteobacterium* that was recently isolated from subsurface sediments (approximately 15 m) at the ERSP-FRC in Oak Ridge, TN. Based upon SSU rRNA gene sequences, the closest cultivable relative is *Anaeromyxobacter dehalogenans* 2CP-C with 96.5% sequence identity. Approximately 105 putative proteins are estimated to contain heme-binding sites, with almost half being multi-heme proteins. This year we also submitted *D. vulgaris* Miyazaki, *D. salixigens*, and *D. desulfuricans* 27774 for sequencing at JGI. We are also preparing *D. termitidis*, *D. fructosivorans*, and two Hanford 100H *Desulfovibrio* sp. for sequencing at JGI.

Artificial communities. Transcriptional and mutational analyses of a coculture composed of *Desulfovibrio vulgaris* Hildenborough and *Methanococcus maripaludis* LL suggests *D. vulgaris* relies upon largely independent mechanisms of energy conservation for syntrophic growth and sulfate-respiration. In particular, the Coo hydrogenase, high-molecular weight cytochrome, [Fe] hydrogenase and [NiFe] hydrogenase play vital roles during syntrophic growth. Additionally, these studies provide a foundation for understanding the more complex communities described below. To better understand the ecology of the syntrophy, we have investigated the effect of genetic diversity on dual culture function by pairing *M. maripaludis* with different *Desulfovibrio* strains. These experiments showed that efficiency of coculture growth is correlated with numerical dominance of *Desulfovibrio* over *M. maripaludis* in co-cultures. Cocultures in which both species were equally predominant or where *M. maripaludis* dominated grew more slowly and achieved lower biomass. To further understand the ecological relevance of these phenotypic differences, we developed a protocol for competing *Desulfovibrio* strains against one another in syntrophic conditions and in monoculture with sulfate as the electron acceptor. To elucidate the potential genetic and phenotypic changes that may predominate in populations growing syntrophically for many generations, we founded 24 independent, but clonally identical populations and have allowed them to evolve for over 200 generations. Preliminary assessments of evolutionary changes indicate that populations evolved higher biomass by 100 generations of evolution, and have evolved increased stability. Currently, we are developing more complex assemblies to examine competitive and cooperative interactions, and factors contributing to community stability. We took advantage of the fully sequenced genomes of *Desulfovibrio vulgaris*, *Geobacter sulfurreducens* PCA, *Geobacter metallireducens* GS-15 and *Methanococcus maripaludis* to engineer various tri-cultures and develop tools to monitor community composition. Initial studies have shown that, tri cultures of *D. vulgaris*,

M. maripaludis and *G. sulfurreducens* growing with fumarate and iron citrate and *G. metallireducens* growing with nitrate and iron citrate produced methane and some amount of hydrogen.

Stress Experiments

High Throughput Biomass Production. Producing large quantities of high quality and defensibly reproducible cells that have been exposed to specific environmental stressors is critical to high throughput and concomitant analyses using transcriptomics, proteomics, metabolomics, and lipidomics. For the past five years, our ESPP project has developed defined media, stock culture handling, scale-up protocols, bioreactors, and cell harvesting protocols to maximize throughput for simultaneous sampling for lipidomics, transcriptomics, proteomics, and metabolomics. In the past five years we have produced biomass for 300 (150 in the last year) integrated experiments (oxygen, NaCl, NO₃, NO₂, heat shock, cold shock, pH, Cr, and mutants Fur, Zur, Per, and MP(-)) each with as much as 400 liters of mid-log phase cells (3 x 10⁸ cells/ml). This year we have also focused on adapting these techniques to *Geobacter metallireducens* and have begun stress biomass production for comparison to *DvH* and *Shewanella*. To determine the optimal growth conditions and determine the minimum inhibitory concentration (MIC) of different stressors we adapted plate reader technology using Biolog and Omnilog readers using anaerobic bags and sealed plates. This has enabled us to link nitrogen utilization plate substrates to specific pathways for *D. vulgaris* as defined by the KEGG database. Characterization of the megaplasmid containing strain was confirmed by the plasmid sequence detailing the nitrogen fixation genes present on the plasmid. Crude mixed culture consisting of *D. vulgaris* and *G. metallireducens* showed PM profiles distinctly different from the profiles of the cultures run solo. We are in the process of modifying techniques to amplify signal of non-SRB organisms with low biomass for further phenotypic profiling.

Phenotypic Responses. We have generated a large set of phenotypic data that suggest analysis of the strain DePue genome sequence will provide important insights into the acquisition of metal-resistance absent in the closely related strain, *D. vulgaris* Hildenborough. An initial phenotypic characterization of a novel *Desulfovibrio* species isolated from the Hanford demonstration site has been completed and DNA is now being prepared for genome sequencing. We have completed extensive phenotypic comparisons of a large study set of *Desulfovibrio* species (14 different strains), as a prelude to continued comparative studies of fitness and evolution. Based upon

* Presenting author

transcriptomic data, a mutant was generated in a gene annotated as a sensory-box protein. Phenotypic analyses indicated that the mutant was deficient in stationary-phase survival. The lack of viability is caused by increases in the sulfide concentration. The data indicated the conserved hypothetical protein was important for sensing sulfide and directly or indirectly regulated a putative set of genes important for the cellular response to end-product accumulation.

Synchrotron FTIR Spectromicroscopy for Real-Time Stress Analysis. Synchrotron FTIR Spectromicroscopy approaches will be discussed in another presentation at this meeting.

38 Applications of Systems Biology Approaches to Understanding Artificial Microbial Consortia and Environmental Communities in the VIMSS Applied Environmental Microbiology Core

Christopher Schadt^{1,4*} (schadtcw@ornl.gov), Zamin Yang,^{1,4} Amudhan Venkateswaran,^{1,4} Meghan Drake,¹ Susan Carroll,^{1,4} Dawn Klingeman,^{1,4} Mircea Podar,^{1,4} Thomas Phelps,^{1,4} Steven Brown,^{1,4} Anthony Palumbo,^{1,4} Sergey Stolyar,^{2,4} Christopher Walker,^{2,4} David Stahl,^{2,4} Terry C. Hazen,^{3,4} and Martin Keller^{1,4}

¹Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; ²Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington; ³Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; and ⁴VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

Cultivation of single species has been at the central core of experimental microbiology for more than a century but offers only a glimpse into the biology of microorganisms in nature. Communities, not individual species, control the process rates that drive key biogeochemical cycles, including the transformation of environmental pollutants of concern to DOE. Thus detailed studies of model consortia and communities that mediate such processes that will allow for experimental manipulation and in-depth analysis of the fundamental biology underlying such

systems are essential for advancing DOE objectives. As part of the GTL Environmental Stress Pathway Project (ESPP) team, we are pursuing two projects to advance these objectives.

Methods development for environmental mRNA analysis.

Current technologies applied to environmental samples for RNA transcriptional profiling include RT-PCR and functional gene microarrays. While, tremendous progress has been made in understanding microbial communities due to emergence of these technologies, they bear significant limitations that prevent their application in a high throughput manner to *de novo* communities. We are developing methods for directly sequencing cDNA from environment samples utilizing new high throughput (HT) sequence analysis technologies. Since 80% or more of total RNA from bacteria is represented by the rRNA pool, it is crucial to first remove those components as thoroughly as possible without adversely impacting mRNA quality, quantity and composition, prior to HT sequence based screening. We have compared three different methods to remove rRNAs and enrich mRNAs of *D. vulgaris* Hildenborough (DvH) samples. The first method utilizes biotin-modified oligos complementary to conserved regions in 16S and 23S rRNA and specific removal by binding to streptavidin-coated magnetic beads. The second uses a commercially available exonuclease that specifically hydrolyzes rRNAs bearing a 5' monophosphate group. The third method uses two rounds of reverse transcription, where rRNAs are first reverse transcribed with multiple universal primers for 16S and 23S rRNAs and subsequently the RNA/DNA hybrids and cDNA are removed by sequential digestion with RNaseH and DNaseI. We have evaluated these three methods alone and in combination using microarray-based analysis of transcription levels. All three methods are able to significantly enrich mRNA from rRNA without introducing significant biases. Microarray analysis revealed significant differences in measured mRNA levels in only 0.5% to 5% of genes across the genome as compared to controls. Comparisons of microarray results with HT sequencing using the Solexa platform are currently ongoing. After further validation, application of these methods could be performed on environmental systems from the Hanford and/or Oak Ridge contaminated sites as part of the VIMSS/ESPP applied environmental core studies, as a complement the DNA based metagenomic analyses already underway at these sites.

Developing manipulatable, laboratory based, higher order microbial consortia. A practical understanding of how community structure leads to process rates and stability is central to DOE objectives in bioremediation and process

* Presenting author

control. Although there are numerous theories relating to stability in macroecology, their relevance to microbial communities is mostly untested. To further these studies we are in the initial stages of assembling model microbial consortia in the laboratory that will allow us to study and manipulate community interactions in a controlled manner and test the stress responses of the assemblages. The model organisms now used by the VIMSS/ESPP team will serve for constructing initial consortia, encompassing sulfate-reducers (*DvH*), iron and uranium reducers (*G. metallireducens*) and methanogens (*M. maripaludis*). The genomes of all these strains have been sequenced, gene expression microarrays are available within the group, and the individual organisms can be genetically manipulated, that will allow unprecedented toolsets to be applied to these controlled communities that would not be possible in natural systems. Additional methanogens, metal-reducing bacteria from DOE contaminated sites (e.g. *Geobacter*, *Anaeromyxobacter*, and *Desulfovibrio* sp.) and heterotrophic *clostridia* that can provide end products of cellulose fermentation (ethanol, acetate and lactate) to the other community members are also to be added to consortia designs. Three member consortia combinations including *C. acetobutylicum*, *DvH* and *G. lovelii*, as well as, additional consortia with *DvH*, *G. metallireducens*, and *M. maripaludis* have been tested thus far in the past few months. Methods for tracking population dynamics of the consortia members such as qPCR and FISH have also been developed and have shown relatively stable assemblages of these species can be achieved. Higher order consortia studies incorporating 4 and more community members are ongoing.

Together these studies will enable us to do in depth studies of stress mechanisms within environmental and model consortia systems, and understand how the detailed mechanisms outlined using pure culture laboratory systems within the Functional Genomics Core of the VIMSS/ESPP project, translate into the relationships and activities observed in more complex constructed consortia as well as ultimately into environmental microbial communities.

39 Growth Rate and Productivity of a Microbial Mutualism Depends on the *Desulfovibrio* Genotype

Kristina L. Hillesland* (hilleskl@u.washington.edu) and David A. Stahl

Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington and VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

In a syntrophic mutualism, two species work together to gain energy from thermodynamically unfavorable reactions. This energy is partitioned between the species, but it is unclear how this occurs and if it is a parameter that can vary depending on the phenotype of each mutualist. In a genetically variable population, genotypes that are capable of acquiring a greater share of the energy from the syntrophic interaction may have a competitive advantage. If such unequal distribution has a negative effect on the overall efficiency of the mutualism, then natural selection could cause instability in these associations. Thus, it is important to understand the relationships between the overall efficiency of syntrophic growth, the allocation of resources among mutualists, and whether these relationships are genetically based. To explore these possibilities, we paired *Methanococcus maripaludis* with several species of the sulfate reducer *Desulfovibrio* and cultured them in the absence of an electron acceptor for *Desulfovibrio*, with lactate as electron donor. In these conditions, *Desulfovibrio* ferments lactate and produces hydrogen that *M. maripaludis* consumes for growth, thereby maintaining thermodynamically favorable conditions for lactate fermentation. We found that growth rate and biomass produced during syntrophic growth varied considerably depending on the genotype of *Desulfovibrio*. To elucidate the relationship between co-culture growth efficiency and composition, we used a study set of four co-cultures that varied in growth rate and biomass yield to track the population dynamics of each species. We estimated the abundance of each species at several time intervals from the relative concentrations of 16S rRNA. Co-cultures containing *D. desulfuricans* 27774 or *D. vulgaris* Hildenborough grew more quickly and achieved a higher OD_{600nm} than co-cultures containing either *D. G20* or *D. vulgaris* oxamicus. In the faster growing co-cultures, *Desulfovibrio* dominated over *Methanococcus* at least 2-fold, but the relative numbers in the slower growing co-cultures were roughly equal or

* Presenting author

avored *Methanococcus*. These results are consistent with results of our flux-balance model of a two species association, which showed that *Desulfovibrio* should dominate 2-2.5-fold over *Methanococcus* in syntrophies that are growing optimally. Thus, unequal allocation of energy towards *Desulfovibrio* may be explained by fluxes through the stoichiometric metabolic network underlying the mutualistic relationship, and may even benefit *Methanococcus* by enabling faster growth.

40 Biodiversity and Spatial Concordance of an in Situ System for Uranium Bioreduction

GTL

C. Hwang,^{1,7} W.-M. Wu,² T.J. Gentry,³ J. Carley,⁴ S.L. Carroll,⁴ D.B. Watson,⁴ P.M. Jardine,⁴ **J. Zhou,**^{5,8} C.S. Criddle,² and **M.W. Fields**^{6,7,8*} (matthew.fields@erc.montana.edu)

¹Department of Microbiology, Miami University, Oxford, Ohio; ²Department of Civil and Environmental Engineering, Stanford University, Stanford, California; ³Department of Crop and Soil Sciences, Texas A&M University, College Station, Texas; ⁴Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; ⁵Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma; ⁶Department of Microbiology and ⁷Center for Biofilm Engineering, Montana State University, Bozeman, Montana; and ⁸Virtual Institute of Microbial Stress and Survival <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

The elucidation of how populations of interest interact in a given community and how the community responds to stress and perturbations can help infer the interplay between stress pathways and gene networks that help optimize bacterial biochemistry. A goal of VIMSS is to characterize the responses of bacterial communities at multiple levels of resolution in order to understand biochemical capacity at DOE waste sites. The current work uses a series of re-circulating wells that create a subsurface bioreactor to stimulate microbial growth for *in situ* U(VI) immobilization (Wu et al. ESandT 41:5716-5723) and the bacterial communities were tracked over time and space in response to nutrient influx. Bacterial community dynamics were investigated in a series of re-circulating wells that created a subsurface “bio-reduction zone” to stimulate bacterial growth with ethanol for in situ bioremediation of U(VI) at the Field

Research Center of the U.S. Department of Energy, Oak Ridge, TN. Different experiments were conducted to alter the subsurface environment to better understand strategies that would improve the remediation process. Within this framework, the interrelationships between the biogeochemistry were studied in order to characterize the community and ecosystem ecology with respect to microbiology of an engineered system. Bacterial community composition and structure of groundwater samples were analyzed via clone libraries of partial SSU rRNA genes. UniFrac analyses showed that the bacterial community in each of the wells developed changes during the bioremediation process, and the changes could be attributed to the variations along temporal and spatial scales. Relationships between community diversity and ecosystem function were idiosyncratic, and these results suggested the population distributions depended on the particular conditions under which the local landscape was investigated. Principal component analysis showed that nitrate, uranium, sulfide, sulfate, and COD were strongly associated with particular bacterial populations. Sequences closely related to nitrate-reducing bacteria were predominant during the initial phase of the remediation process, but sequences representative of sulfate-reducers (*Desulfovibrio* and *Desulfosporosinus spp.*) and metal-reducers (*Geobacter spp.*) were detected at higher levels as uranium levels declined. Ultimately, sequences associated with sulfate-reducing populations predominated. Uranium levels declined below EPA drinking water standards, and community composition and structure were similar in both treatment wells after approximately 1.5 y despite going through different transitions. In addition, when engineering controls were compared to the community structure and composition via canonical ordinations, population distributions could be related with dissolved oxygen control and the presence of bio-stimulant. During the bio-stimulation, population distributions followed geochemical parameters, and these results indicated that bacteria exhibited distributions at the landscape scale in concordance with predictable geochemical factors. The data indicated that relationships between community structure and ecosystem function were idiosyncratic, but temporal and spatial concordance were eventually observed for the two bio-stimulated wells. The strong associations between particular environmental variables and certain population distributions will provide insights into establishing practical and successful remediation strategies in radionuclide-contaminated environments with respect to engineering controls and ecosystem function.

* Presenting author

41

GTL

Studying Rules Governing Microbial Communities

Sergey Stolyar* (sstolyar@u.washington.edu) and David A. Stahl

Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington and Virtual Institute for Microbial Stress and Survival (<http://vimss.lbl.gov>)

Project Goals: See goals for abstract 36.

Understanding the behavior of biological communities presents tremendous challenges because of the complex network of diverse interactions among species. A class of communities of particular interest from ecological, geological, and engineering perspectives is represented by microbial communities that thrive in oxygen-free (anoxic) environments. These communities are vital components in numerous environments including freshwater sediments, the subsurface, guts of insects and animals, wastewater treatment plants and heavy-metal contaminated environments of concern to DOE. They therefore play a significant role in global cycling of carbon and other biogenic elements. Our prior research established the feasibility of working with a simple two-tier food web composed of two species, each species occupying a distinct trophic position: *Desulfovibrio vulgaris* syntrophically coupled to the hydrogen consuming *Methanococcus maripaludis*. As part of phase two of the Environmental Stress Pathway Project (ESPP) we are now developing more complex assemblies to examine competitive and cooperative interactions, and factors contributing to community stability. We are using organisms having fully sequenced genomes (*Desulfovibrio vulgaris*, *Geobacter sulfurreducens* PCA, *Geobacter metallireducens* GS-15 and *Methanococcus maripaludis*) to construct different tri-cultures and develop tools to monitor community composition. *G. sulfurreducens* and *G. metallireducens* consume acetate and use alternative electron acceptors, including nitrate, fumarate and iron. The ecological relevance of this electron acceptor versatility in constructed microbial food webs is being examined by measuring growth kinetics and yields in relationship to the modynamic predictions. Initial studies have shown that lactate-grown tri-cultures of *D. vulgaris* and *M. maripaludis* combined with either *G. sulfurreducens* (plus fumarate or iron citrate) or *G. metallireducens* (plus nitrate or iron citrate) evolved both methane and hydrogen at different stages of growth in batch culture. Tri-cultures with *G. sulfurreducens* growing with fumarate and tri culture with *G. metallireducens*

* Presenting author

growing with nitrate demonstrated comparable growth rates and biomass yields despite differences in predicted total free energy: -671 kJ for fumarate and -1183 kJ for nitrate.

42

GTL

Energy Conservation in a Biogeochemically Significant Microbial Mutualism

Christopher B. Walker,^{1,8,*} Zhili He,^{2,3,8} Zamin K. Yang,^{2,8} Janet Jacobsen,^{4,8} Joseph A. Ringbauer, Jr.,^{5,8} Qiang He,^{2,8} Jizhong Zhou,^{2,3,8} Gerrit Voordouw,⁶ Judy D. Wall,^{5,8} Adam P. Arkin^{4,7,8} (aparkin@lbl.gov), Terry C. Hazen,^{4,8} Sergey Stolyar,^{1,8} and David A. Stahl^{1,8}

¹Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington; ²Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; ³Institute For Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; ⁴Lawrence Berkeley National Laboratory, Berkeley, California; ⁵University of Missouri, Columbia, Missouri; ⁶University of Calgary, Calgary, Alberta, Canada; ⁷Department of Bioengineering, University of California, Berkeley, California; and ⁸Virtual Institute for Microbial Stress and Survival (<http://vimss.lbl.gov>)

Project Goals: See goals for abstract 36.

Complex interactions govern microbial communities in both pristine and contaminated environments. Unfortunately, a limited understanding exists regarding the interactions sustaining these communities. Without a deeper knowledge of the molecular basis driving the community structure and function, bioremediation of contaminated sites suffers from inefficient or ineffective design strategies. The VIMSS/ESPP2 project aims at resolving the molecular basis for microbial organisms and communities found at heavy-metal contaminated environments, such as the Hanford 100H site. Sulfate-reducing microbes (SRM) commonly compose significant fractions of the microbial community in contaminated anaerobic sites. The VIMSS/ESPP2 project extensively examined a representative SRM, *Desulfovibrio vulgaris* Hildenborough, building a detailed understanding of stress response mechanisms through intensive monoculture study. However, *D. vulgaris* often populates environments deficient in sulfate, relying upon syntrophic associations with

hydrogenotrophic methanogens for continued growth. Investigation of an archetypical community composed of *D. vulgaris* Hildenborough and a representative methanogen, *Methanococcus maripaludis* LL, serves as a basis for understanding the physiological differences between growth modalities. Using transcriptional analysis, we demonstrate that continuously grown cultures of *D. vulgaris* Hildenborough, up-regulate a broad suite of electron transfer enzymes during syntrophic growth. Mutational analyses indicate key roles for four enzymes (Coo, Hmc, Hyd and Hyn) not essential for sulfate-respiration. Specifically, these results provide a developed molecular basis for understanding this “community of two” while also serving as the foundation for future VIMSS/ESPP2 community analysis. More generally, these results suggest syntrophic growth and sulfate-respiration rely upon largely independent energy generation pathways.

43

ESPP Functional Genomics and Imaging Core: Cell Wide Analysis of Metal-Reducing Bacteria

Aindrila Mukhopadhyay^{1,6*} Edward Baidoo,^{1,6} **Kelly Bender**^{5,6} (bender@micro.siu.edu), Peter Benke,^{1,6} Swapnil Chhabra,^{1,6} Gabriela Chirica,^{2,6} Elliot Drury,^{3,6} Dwayne Elias,^{3,6} Veronica Fok,^{1,6} Sara Gaucher,^{2,6} Masood Hadi,^{2,6} Zhili He,^{4,6} Chris Hemme,^{4,6} **Jay Keasling**^{1,6} (keasling@berkeley.edu), Kimberly Keller,^{3,6} Yooli Light,^{2,6} Eric Luning,^{1,6} Robert Meagher,^{2,6} Francesco Pingitore,^{1,6} Alyssa Redding,^{1,6} Jarrod Robertson,^{3,6} Rajat Sapra,^{2,6} Matthew Shirley,^{3,6} **Anup Singh**^{2,6} (aksingh@sandia.gov), Yinjie Tang,^{1,6} **Judy Wall**^{3,6} (wallj@missouri.edu), Huei-Che Yen,^{3,6} Grant Zane,^{3,6} Aifen Zhou,^{4,6} and **Jizhong Zhou**^{4,6} (jzhou@rccc.ou.edu)

¹Lawrence Berkeley National Laboratory, Berkeley, California; ²Sandia National Laboratories, Livermore, California; ³University of Missouri, Columbia, Missouri; ⁴University of Oklahoma, Norman, Oklahoma; ⁵Southern Illinois University; and ⁶Virtual Institute of Microbial Stress and Survival <http://vimss.lbl.gov>

Project Goals: See goals for abstract 36.

A fundamental goal of the Environmental Stress Pathway Project (ESPP) of the Virtual Institute of Stress and survival (VIMSS) is a rigorous understanding of *Desulfovibrio vulgaris* Hildenborough physiology and its ability to survive in its environment. Such knowledge

will be critical in discerning the biogeochemistry at metal contaminated sites, for bioremediation and natural attenuation for toxic metals. The Functional Genomics and Imaging Core (FGIC) focuses on the measurement of these responses at a cell wide level using systems biology approaches.

Progress in the last one year built upon our optimized pipeline for generating biomass for various functional genomics studies and utilized improved genetic methods. Numerous additional transcriptomics data sets were added to our compendium of stress response studies. These included peroxide stress, low oxygen stress, high and low pH stress and alteration of growth conditions (e.g presence of methionine, alternate electron donors etc). To understand how genotype and environment interact to determine the phenotype and fitness of an organism, a long-term evolution experiment was also conducted to examine the dynamics and adaptation of *D. vulgaris* under extended salt exposure. For many of these stresses iTRAQ based quantitative proteomics and CE-MS based metabolite studies were also conducted. Improved genetic methods were employed to create several critical knock out mutants (e.g. *echA*, *qmoABC* and *tatA*) and several were characterized via growth and transcriptomics studies. Progress was also made in extending transcriptomics analysis to examine alternate *D. vulgaris* physiological states such as in biofilms and growth in syntrophic co-culture with *Methanococcus maripaludis*. Methods to conduct iTRAQ proteomics and stable isotopomer (¹³C) based metabolic flux analysis were also developed for studying co-cultures. A novel FTICR-MS based method for a comparative ¹²C/¹³C based metabolite analysis is being developed and will enable a direct comparison of control cultures to experimental samples. Additionally we continued to collect cell wide data in *Shewanella oneidensis* and *Geobacter metallireducens* for comparative studies. Great progress was made in improving extraction and high throughput of metabolite studies. Metabolite extraction and CE-MS detection for several hundred metabolites can now be conducted for these non-model organisms using high resolution separation and high resolution mass spectrometric methods.

Continued studies to map cell wide responses have also emphasized the importance of changes that require orthologous measurements. With this in mind a novel protocol to monitor protein-protein interactions and redox state of the proteins has been developed. In an effort to supplement model development and elucidate intricacies of stress response cascades, comprehensive methods for identifying alternative regulatory mechanisms such as small non-coding RNAs are also underway. To optimize the use of the large amounts of data being

* Presenting author

collected, several data mining efforts were initiated. For example, iTRAQ data sets from the multiple stress response studies were mined for potential post translational modifications and confirmation of hypothetical proteins while ^{13}C flux data were used to confirm gene annotation and assess missing steps in metabolic pathways. Work in underway in collaboration with the computational core to set up searchable databases of our proteomics and metabolite data also.

44

GTL

VIMSS ESPP: Deciphering the Roles of Two-Component Systems in *Desulfovibrio vulgaris* Hildenborough

Aindrila Mukhopadhyay^{1,2*} (AMukhopadhyay@lbl.gov), Dominique Joyner,^{1,3} Eric Luning,^{1,2} Kimberly Keller,^{1,4} Jarrod Robertson,^{1,4} Grant Zane,^{1,4} Janet Jacobsen,^{1,2} Morgan Price,^{1,2} Swapnil Chhabra,^{1,2} **Terry C. Hazen**,^{1,3} **Adam P. Arkin**,^{1,2} **Judy Wall**,^{1,4} and **Jay Keasling**^{1,2}

¹Virtual Institute of Microbial Stress and Survival; <http://vimss.lbl.gov>; ²Physical Biosciences Division and ³Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; and ⁴Biochemistry and Molecular Microbiology and Immunology Departments, University of Missouri, Columbia, Missouri

Project Goals: See goals for abstract 36.

One of the primary goals of the Environmental Stress Pathway Project (ESPP) is to map the response of the anaerobic sulfate reducing soil bacterium, *Desulfovibrio vulgaris* Hildenborough to its environment. Two component systems, comprised of Histidine Kinase and Response regulator proteins, present the primary and ubiquitous mechanism in bacteria for initiating cellular response towards a wide variety of environmental conditions. In *D. vulgaris* Hildenborough, more than 70 such systems have been predicted, but remain mostly uncharacterized. The ability of *D. vulgaris* to survive in its environment is no doubt linked with the activity of genes modulated by these two component signal transduction systems. To map the two component systems to the genes they modulate, the availability of deletion mutants provides an important tool. Here we present an overview of the predicted histidine kinases in *D. vulgaris* and describe a strategy to create library of histidine kinase knock out mutants in *D. vulgaris*. We use the OmniLog[®] workflow to conduct a wide phenotypic characterization

of the knock out mutants generated. To illustrate our strategy we present results from our study of the histidine kinase in the predicted *kdp* operon of *D. vulgaris*. The high-affinity potassium uptake Kdp complex is well characterized in other bacteria where it facilitates K^+ uptake in low K^+ or high Na^+ conditions. Typically, the activity of the Kdp system is modulated by the KdpD/E two-component signal transduction system, where KdpD is the sensor histidine kinase and KdpE is the response regulator. The *D. vulgaris kdp* operon contains a gene with predicted response regulator function and two separate genes annotated for the sensor kinase function (*kdpD* and DVU3335). Interestingly, only one of these two, DVU3335, contains a conserved histidine kinase domain which is absent the *D. vulgaris kdpD* candidate. However, DVU3335 does not encode the well-conserved motifs associated with KdpD. We created a knock out mutant in the DVU3335 gene. The DVU3335 knock out strain showed a growth deficiency in low K^+ conditions and when exposed to low K^+ conditions was unable to upregulate genes in the *kdp* operon. Phenotypic microarrays were used to obtain a broader comparison of the mutant and wild type strains. Our results show that the major differences between the wild type and the mutant are in response to salt stress and support the role of DVU3335 in modulating K^+ uptake during low K^+ and high Na^+ conditions.

45

GTL

Observing Polyglucose Metabolism and Transient Oxygen Stress in Obligate Anaerobes in Vivo

Hoi-Ying N. Holman^{1,2*} (hyholman@lbl.gov), Eleanor Wozel,¹ Zhang Lin,^{1,3} Luis Comolli,¹ Kenneth H. Downing,¹ **Matthew Fields**,^{2,4} and **Terry C. Hazen**^{1,2}

¹Lawrence Berkeley National Laboratory, Berkeley, California; ²Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov>; ³Fujian Institute of Research on Structure of Matter, Chinese Academy of Science, China; and ⁴Montana State University, Bozeman, Montana

Project Goals: See goals for abstract 36.

Our purpose is to provide the VIMSS/ESPP(1) project with molecular information which will improve the existing understanding of microbial adaptive response strategies, and also will enable a better management and modeling of microbes in subsurface environments for cleanup of hazardous waste sites. Specifically, we studied

* Presenting author

at a chemical level how the obligate anaerobic *Desulfovibrio vulgaris* could survive in a variety of environments which periodically become aerobic.

Aerobic respiration of intracellular polyglucose reserves is postulated to play a central role in oxygen adaptive response in obligatory anaerobes like *Desulfovibrios*, but it has been difficult to probe this event at chemical scale *in vivo*. Here we presented a non-invasive synchrotron infrared (SIR) spectromicroscopy approach to reveal time-dependent composition and structure changes at a lateral scale of several individual *D. vulgaris*. The advantage of infrared spectroscopy is that it is non-invasive, and it uses vibration movements of atoms and chemical bonds within functional groups of biomolecules as an intrinsic contrasting mechanism; thus it allows one to immediately detect composition and structure changes within cells. The advantage of using a synchrotron light source is that its high brightness allows us to detect signals ~1000 times weaker than the conventional infrared spectroscopy allows us to detect.

Comparative analysis of SIR spectra of the same individual *D. vulgaris* exposed to air-level oxygen at different time points reveals chronological information regarding the level of oxidative stress and the extent of cellular injury and repair. These results, together with high-resolution microscopy images, mark a critical step toward the use of SIR spectromicroscopy as an uninterrupted microprobe at a chemical scale level of physiological events in microbiology applications.

46 Phenotypic Characterization of Microorganisms by Barcoded Transposon Mutagenesis

—
GTL

Adam Deutschbauer,^{1,4} Julia Oh,² Morgan Price,^{1,4} Paramvir Dehal^{1,4*} (PSDehal@lbl.gov), Dan Bruno,² Marlene Henriquez,^{1,4} Romy Chakraborty,^{1,4} **Terry C. Hazen**,^{1,4} Corey Nislow,³ Guri Giaever,³ Ronald W. Davis,² and **Adam P. Arkin**^{1,4}

¹Lawrence Berkeley National Laboratory, Berkeley, California; ²Stanford Genome Technology Center, Stanford, California; ³University of Toronto, Toronto, Ontario, Canada; and ⁴VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

The Environmental Stress Pathway Project (ESPP) aims to elucidate the molecular mechanisms by which micro-

bial communities affect stress response and activity in the field and in the laboratory using sulfate reducing bacteria (SRB) as a model. To accomplish our goal of linking field observations to those in the laboratory, a systems-level understanding of SRB genome function is necessary. To meet this challenge, we are developing a mutagenesis and phenotyping strategy that is comprehensive across the genome and applicable to any microorganism amenable to transposon mutagenesis. Here we describe the application of our strategy to *Shewanella oneidensis* MR1 and the SRB *Desulfovibrio desulfuricans* G20. We have cloned and sequence-verified ~3000 barcode modules into a Gateway entry vector. Each module is a 175 base pair element containing two unique 20 base pair sequences, the UPTAG and DOWNTAG, flanked by common PCR priming sites. Each module can then be rapidly transferred *in vitro* to any DNA element, such as a transposon, that is made Gateway compatible. Transposon mutants marked by the modules will be sequenced to determine which of the ~3000 barcode modules was used and which gene was disrupted. Transposon mutants can be rapidly re-arrayed into a single pool containing ~3000 uniquely tagged, sequence-verified mutant strains. By sequencing saturating numbers of transposon mutants, we can identify and assay mutants in most nonessential genes in a given genome. The fitness of each mutant in the pool will be monitored in parallel by the hybridization of the barcodes to an Affymetrix microarray containing the barcode complements in a system identical to that used for the yeast deletion collection. Compared to other approaches for the parallel analysis of transposon mutants such as signature tagged mutagenesis, genetic footprinting, and transposon site hybridization, our approach offers much higher throughput, a single microarray design is universal for any organism, single mutational events are assayed, and mutant strains are archived for verification, distribution, and the systematic genetic interrogation of individual pathways.

The successful completion of this project will enable the quantitative phenotypic analysis of thousands of mutants across a wide range of conditions. These data will be used to assign gene function on a global scale, aid in the identification of missing metabolic enzymes, and provide insight into the functional connectivity of different pathways. Our results in *Desulfovibrio desulfuricans* G20 will be extrapolated *via* comparative genomic analysis to other sequenced SRBs including *Desulfovibrio vulgaris* Hildenborough. Consequently, our findings will aid in the interpretation of both laboratory and field data collected by the ESPP.

* Presenting author

47

The Development of a Markerless Deletion System in *Desulfovibrio vulgaris* Hildenborough

Kimberly L. Keller^{1,3*} (kellerkl@missouri.edu), **Kelly S. Bender**,^{2,3} and **Judy D. Wall**^{1,3} (WallJ@missouri.edu)

¹Biochemistry Department, University of Missouri, Columbia, Missouri; ²Microbiology Department, Southern Illinois University, Carbondale, Illinois; and ³VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

To fully explore microbial community dynamics, stability of the composition with time and changing nutritional and environmental factors must be explored. In order to confirm sources and sinks of metabolites, both during degradation and biosynthesis, it would be most useful to create pivotal deletions in various members of the community. We are pursuing genetic tools that can possibly be applied to strains with limited genetic accessibility. These tools are being developed in *Desulfovibrio* and include an inframe deletion procedure and plasmid modification in extracts to facilitate genetic exchange processes.

To properly study metabolic pathways, it is necessary to delete several genes that may have compensatory activities. Our model system, the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough has seen enormous progress in genetic manipulation; however, the current deletion method of marker exchange mutagenesis does not allow for easy selection of multiple sequential gene deletions because of the low number of selectable markers available in *D. vulgaris*. To broaden the repertoire of genetic tools available for manipulation in *D. vulgaris*, an in-frame markerless deletion system is being developed based on the *upp*-encoded uracil phosphoribosyltransferase as an element for a counterselection strategy. In wild-type *D. vulgaris*, growth is inhibited by the toxic pyrimidine analog 5-fluorouracil (5-FU), whereas a mutant bearing a deletion of the *upp* gene is resistant to 5-FU. The introduction of a plasmid containing the wild-type *upp* gene expressed constitutively from the *apb(5')*-III promoter (the promoter for the kanamycin resistance gene in Tn5) into the *upp* deletion strain restored sensitivity to 5-FU. This observation is the basis for the establishment of a two-step integration and excision strategy for the deletion of genes of interest. Since this deletion does not contain an antibiotic cassette,

GTL

multiple gene deletions can be generated in a single strain using this method.

This in-frame markerless deletion method is currently being evaluated through the construction of a deletion of the putative formate dehydrogenase alpha- and beta- subunits, DVU0587 and DVU0588. In addition, Gateway Technology methods are being developed that would expedite the process of generating the required deletion vectors by the construction of a destination vector containing the constitutively expressed wild-type *upp* gene. This new method is being utilized to generate a deletion for the R-subunit (DVU1703) of a type I restriction-modification system.

48

Expression Profiling and Gene Association of Hypothetical and Conserved Hypothetical Genes in *Desulfovibrio vulgaris* Leads to Functional Annotation

Dwayne A. Elias,^{1,5*} Elliot C. Drury,^{1,5} Alyssa M. Redding,^{2,5} **Aindrila Mukhopadhyay**,^{2,5} Huei-Che B. Yen,^{1,5} Katherine H. Huang,^{3,5} **Terry C. Hazen**,^{2,5} **Adam P. Arkin**^{2,4,5} (aparkin@lbl.gov), and **Judy D. Wall**^{1,5}

¹Department of Biochemistry, University of Missouri, Columbia, Missouri; ²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; ³Massachusetts Institute of Technology, Cambridge, Massachusetts; ⁴Department of Bioengineering, University of California, Berkeley, California; and ⁵Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov>

Project Goals: See goals for abstract 36.

The annotation of predicted open reading frames within sequenced bacterial genomes has derived from physiological experimentation of pure bacterial cultures. This has resulted in a large residual fraction of unannotated genes. As studies move to more and more environmentally relevant microbial communities, these hypothetical and conserved hypothetical genes may well prove to encode critical functions. We have to date observed numerous hypothetical and conserved genes that respond to a single stress or condition, including in a sulfate-reducing bacteria (SRB)/methanogen co-culture. It is expected that such information will be useful in

GTL

* Presenting author

the shorter-term ESPP2 project goal of sequencing and annotating additional SRB from DOE sites, as well as the longer-term project goals of understanding the role of SRB in multi-organism and syntrophic cultures, including targeting genes for deletion that may be important in such conditions.

Hypothetical and conserved hypothetical genes consistently make up 30% or more of sequenced bacterial genomes, with few reports confirming their expression at either the rRNA or protein level. This is the case for the SRB *Desulfovibrio vulgaris* Hildenborough where 348 of the total 3534 genes are currently annotated as conserved or conserved hypothetical (9.7%) with an additional 889 hypothetical genes (25.0%). Given this large complement of the genome, it is plausible that some of these genes serve significant cellular functions that may well range from regulation to presently unknown steps in carbon or electron flux. The goals of this study were to determine which of these genes actually encodes a protein and then to infer a more functionally based annotation. In order to accomplish this, we have compiled and explored the microarray and MS-based iTRAC proteomic expression profiles for the expected 1237 hypothetical and conserved hypothetical proteins in *D. vulgaris* from the ten environmental stresses for which data have been collected in the VIMSS/ESPP project. Categories for examining the expression data included the predicted operon arrangements (monocistronic versus polycistronic operons), the basal transcription levels, and any differential expression in stressed cells. Overall, we are presently able to confirm the expression of 1219 genes at the mRNA level and 265 genes at both the mRNA and protein level, while there was no evidence for either mRNA or protein detectable for 17 genes. While the number of microarray studies outweighs proteomics, the abundance values indicating differential expression at the protein level were consistent with the microarray results when data were available. The monocistronic genes were then reannotated based on several factors including their stress response profile and COG information when available, while polycistronic genes were reannotated using the expression profiles, COG information and association with better annotated genes within the operon. Additionally, analyses of deletion mutants of *fur* and *perR*, encoding global regulatory systems revealed a number of genes apparently regulated directly or indirectly by Fur and/or PerR.

The validity of such inferred assignments can only be ascertained by interruption or deletion of the gene with further analysis. To this end, two targeted deletion mutants were constructed and used as test cases to determine the accuracy of the revised functional annotations. In each case the phenotype obtained was in agreement

with the expanded annotation. We are also in the process of testing several mutants that have been isolated from a random transposon library. Through these efforts, a more precise inter-organism comparison will be possible, thereby yielding better clues as to the function of conserved or hypothetical genes in other organisms. Finally, some of the physiological and metabolic aspects of *D. vulgaris* that are currently not understood may be solved thus aiding bioremediative efforts by understanding the mechanisms by which this bacterium survives the stresses and community effects likely experienced at DOE contaminated sites.

49

GTL

Identification of a Small Non-Coding RNA in *Desulfovibrio vulgaris*

Kelly S. Bender^{2,3*} (bender@micro.siu.edu) and
Judy D. Wall^{1,3}

¹Biochemistry Department, University of Missouri, Columbia, Missouri; ²Microbiology Department, Southern Illinois University, Carbondale, Illinois; and ³VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

Because one of the central aims of the 'Environmental Stress Pathway Project' is to elucidate regulatory networks critical to processes of interest to the DOE, the Computational Core has garnered valuable transcriptional and proteomic profiles under various environmental stress conditions. This data has been essential to enhancing our biological systems knowledge of the model metal reducer *Desulfovibrio vulgaris*. To further understand how this organism and its relatives compete in the environment and regulate their metabolism in contaminated sites, additional studies on intricate regulatory cascades are imperative. One potential alternative regulatory mechanism currently being targeted by the ESPP is that of small non-coding RNA molecules (sRNAs). Ranging in size from 20–200 nucleotides (nt), sRNAs predominantly affect gene regulation by binding to complementary mRNA in an anti-sense fashion and therefore provide an immediate regulatory response independent of protein modification. Whereas data are available for sRNAs in such prokaryotes as *Escherichia coli*, *Archaeoglobus fulgidus*, *Pseudomonas aeruginosa*, and *Vibrio* species, no information is available on metal-reducing bacteria, or for that matter, members of the delta-*proteobacteria*.

* Presenting author

In an effort to identify sRNAs in *D. vulgaris*, a strategy for cloning total RNA ranging in size from 20–200 nt was employed. Following addition of directional aptamer sequences, cDNAs were produced and cloned for sequencing. Sequence analysis of a small portion of the resulting cDNA library yielded two identical ~65 nt sRNA clones (Dv-sRNA2) possessing complementary sequence to the RBS of open reading frame (ORF) DVU0678. While DVU0678 is adjacent to the Dv-sRNA2 gene, the ORF is transcribed from the opposite chromosomal strand. Northern analysis specialized for sRNAs verified the expression of Dv-sRNA2 as an individual transcript under anaerobic lactate/sulfate growth (LS4D medium). These data suggest that when Dv-sRNA2 is transcribed, translation of DVU0678 will be inhibited. DVU0678 has been annotated to encode a putative 34 amino acid protein unique to *D. vulgaris* strains Hildenborough and DP4, hampering our abilities to discern the role of DVU0678 in the cell. Further sequence analysis of the Dv-sRNA DNA locus by ‘PromScan’ identified a putative sigma⁵⁴-recognition site (97% probability) 43 nt upstream of the predicted sRNA transcriptional start site and therefore suggests that Dv-sRNA2 may be member of the sigma⁵⁴ regulon. A perfect stem-loop terminator was also identified 26 nt downstream of the Dv-sRNA2 DNA sequence. Current analysis is underway to ascertain the expression profile for this sRNA as well as the effect over-expression has on the physiology and transcriptional response of *D. vulgaris* under multiple environmental conditions.

50 GTL The Dynamics and Genetic Adaptation to Salt Stress in Experimental Evolution of *Desulfovibrio Vulgaris* Hildenborough

Aifen Zhou,^{1,6*} Zhili He,^{1,6} Marcin P. Joachimiak,^{2,6} Paramvir S. Dehal,^{2,6} Adam P. Arkin,^{2,6} Kristina Hillesland,^{4,6} David Stahl,^{4,6} Judy Wall,^{5,6} Terry C. Hazen,^{3,6} and Jizhong Zhou^{1,6} (jzhou@ou.edu)

¹Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; ²Physical Biosciences Division and ³Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, California; ⁴Civil and Environmental Engineering, University of Washington, Seattle, Washington; ⁵Departments of Biochemistry and Molecular Microbiology and Immunology,

University of Missouri, Columbia, Missouri; and ⁶Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov>

Project Goals: See goals for abstract 36.

One of the greatest challenges in biology is to understand how genotype and environment interact to determine the phenotype and fitness of an organism. With the recent advances in genome sequencing and high-throughput genomic technologies, it becomes capable to link sub-cellular molecular/metabolic processes with the population-level processes, functions and evolution. One of our goals of the new proposal is to bring the environmental microbe, *D. vulgaris* Hildenborough to the model organism status (Aim 1). This study particularly intends to mimic the environmental conditions (salt stress) to address the evolution of DvH under such conditions in the lab. Such a study is expected to generate different DvH strains, and allows us to identify multiple beneficial mutations for salt adaptation. Therefore, this study directly links stress responses to the evolution of DvH, and will provide information for our integrative understanding of gene function, regulation, networking and evolution of DvH. To determine the long-term evolutionary responses, diversifications and adaptation of DvH to environmental stresses, the control and evolved cell lines (6 lines each) were obtained from a single DvH colony directly from the original glycerol stock. LS4D was used as standard culture medium for the control lines. Evolved lines were cultured on LS4D + 100 mM NaCl. Cells were kept at 37°C and transferred every 48 hrs with one to one hundred dilutions. The cells from every 100 generations were archived. The results demonstrated that the adaptation of DvH to salt stress was a dynamical process. The enhanced salt tolerance to higher salt (LS4D + 250 mM NaCl) of evolved lines was observed at 300 generations; and this phenomenon became more and more obvious with the increase of generations. Compared to the ancestor and paralleled control lines, both the growth rate and final biomass of the evolved lines were higher. The de-adaptation experiment on 1000 generation evolved lines provided the evidence that the phenotype was due to the genetic change instead of physiological adaptation. The gene expression profile of the 1000 generation evolved lines showed that some poly-cistronic operons such as hmcF-E-D-C-B-A (functional genes), rrf2-rrf1 (regulatory genes), LysA-2-LysX (functional genes) and DVU3290-3291-3292 (glutamate synthase) were significantly up-regulated compared to the control lines. Next, de-adaptation experiment need to be done on different generation samples to confirm that the beneficial genetic mutations are stable and genome sequencing will be performed to reveal the possible genetic mutations.

* Presenting author

Applications of GeoChip to Examine Functional Microbial Communities in Metal Contaminated Environments

Joy D. Van Nostrand,^{1,8*} Liyou Wu,¹ Sanghoon Kang,¹ Patricia Waldron,¹ Yuting Liang,¹ Ye Deng,¹ Zhili He,^{1,8} Weimin Wu,² Heidi Gough,³ Sue Carroll,⁴ Chris Schadt,^{4,8} Anthony Palumbo,^{4,8} Dave Watson,⁴ Craig Criddle,² Phil Jardine,⁴ Brett Baldwin,⁵ Aaron Peacock,⁵ Phil Long,⁶ David Stahl,^{3,8} **Terry C. Hazen,**^{7,8} and **Jizhong Z. Zhou**^{1,8} (jzhou@ou.edu)

¹University of Oklahoma, Norman, Oklahoma;

²Stanford University, Stanford, California; ³University of Washington, Seattle, Washington; ⁴Oak Ridge National Laboratory, Oak Ridge, Tennessee;

⁵University of Tennessee, Knoxville, Tennessee;

⁶Pacific Northwest National Laboratory, Richland, Washington; ⁷Lawrence Berkeley National Laboratory, Berkeley, California; and ⁸VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

The GeoChip 2.0, a functional gene microarray, allows for the simultaneous detection of >10,000 genes involved in the geochemical cycling of C, N, and S, metal reduction and resistance, and organic contaminant degradation. The GeoChip has been used to examine dynamic functional and structural changes in microbial communities from many different environments. Here, five examples of studies utilizing the GeoChip to examine microbial communities at metal contaminated sites are presented. These studies specifically address and support Aim 3's goal to 'examine the natural community dynamics in the field for correlation to laboratory observations'. These initial studies illustrate the ability of the GeoChip to provide direct linkages between microbial genes/populations and ecosystem processes and functions. The first three studies examined areas within the U.S. DOE's Field Research Center (FRC) in Oak Ridge, TN. (1) Microbial communities within a pilot-scale test system established for the biostimulation of U(VI) reduction in the subsurface by injection of ethanol were examined. The microarray data indicated that during the U(VI) reduction period, both FeRB and SRB populations reached their highest levels at Day 212, followed by a gradual decrease over the following 500 days. The U concentrations in the groundwater were significantly correlated with the total abundance of c-type cytochrome genes and with the total abundance of *dsrAB* (dissimilatory sulfite reductase) genes. Mantel test analysis of microarray and chemical data indicated a

significant correlation between the U concentration and total c-cytochrome or *dsrAB* gene abundance. Changes in more than a dozen individual c-type cytochrome genes and more than 10 *dsrAB*-containing populations showed significant correlations to the changes in U concentration among different time points, indicating their importance in uranium reduction. (2) In a different study of the same system, the effects of dissolved oxygen (DO) and ethanol on the stability of the bioreduced area were examined. Canonical correspondence analysis (CCA) and Mantel test analysis revealed that ethanol and sulfide concentrations showed the greatest correlation to the functional community structure. Detrended correspondence analysis (DCA) showed a shift towards a different community structure after ethanol injections resumed compared to the periods of starvation and exposure to DO. Changes in the functional community structure were similar in both wells; however, the community in FW101-2 was more affected by DO than in FW102-3. This is most likely because FW101-2, located closest to the injection wells, had a greater increase in DO than FW102-3, located further from the injection well. Hierarchical clustering showed that cytochrome c genes grouped based on DO exposure, starvation, or ethanol addition, while dissimilatory sulfite reductase (*dsr*) genes grouped only by starvation or ethanol addition. However, when DO levels increased, the relative abundance of *dsr* genes decreased while cytochrome genes seemed unaffected. Overall, results indicated that ethanol was the main factor affecting community structure, although some changes could be attributed to DO. (3) In the third study from the FRC, analysis of groundwater monitoring wells along a contamination gradient revealed less overlap between wells with different levels of U and NO₃ contamination. While diversity of nitrate-fixation genes decreased in NO₃-contaminated wells, the diversity of metal reduction and resistance genes did not correlate with metal concentrations. Signal intensity did, however, increase in heavily contaminated wells, indicating a larger percentage of organisms with metal-related genes. Sulfate-reduction genes had greater diversity and greater signal intensity in more contaminated wells. Individual principle component analyses (PCA) of the gene diversity and geochemistry of the wells separated them in similar ways. CCA indicated that pH was an important variable that correlated with gene diversity in the lowest-contamination well, while NO₃ and U correlated with the most highly contaminated well. Overall, contaminant level appears to have significant effects on the functional gene diversity along the contaminant plume at the FRC. (4) We have also used GeoChip to examine a Uranium Mill Tailings Remedial Action (UMTRA) site (Rifle, CO). Two adjacent mini-galleries were driven to Fe-reducing and

* Presenting author

SO₄-reducing conditions, respectively, in order to better understand the long-lived U (VI) loss and to constrain the relative impacts of SO₄ and Fe (III) reduction. Cluster analysis results showed samples in the same locations grouped together, regardless of geochemistry. The *dsr* genes increased when conditions were driven to sulfate-reduction. DCAs of both the functional community structure and environmental conditions (Fe²⁺, H₂S, DO, pH, conductivity, potential and Eh) showed background, Fe-reducing, and SO₄-reducing samples clustered together, respectively. CCA of environmental parameters and functional genes indicated Fe²⁺ was the most significant geochemical variable for community structure. (5) Additionally, metal contaminated freshwater lake sediments (Lake DePue, IL) were analyzed to examine the link between functional genes and the environmental gradient of metal contamination. Based on non-metric multidimensional scaling (NMDS), the microbial communities were separated based on sampling regions. In addition, there were different groupings between samples with the highest levels of contamination and the lesser contaminated samples. These studies demonstrate the analytical power of the GeoChip in examining microbial communities. This is the first comprehensive microarray available for studying the functional and biogeochemical cycling potential of microbial communities.

52 Further Development and Applications of GeoChip 3.0 for Microbial Community Analysis

GTL

Zhili He,^{1,5*} Ye Deng,^{1,5} Joy D. Van Nostrand,^{1,5} Liyou Wu,^{1,5} Christopher Hemme,^{1,5} Terry J. Gentry,² Adam P. Arkin,^{3,5} Terry C. Hazen,^{4,5} and Jizhong Zhou^{1,5} (jzhou@ou.edu)

¹University of Oklahoma, Norman, Oklahoma; ²Texas A&M University, College Station, Texas; ³Physical BioScience Division, Lawrence Berkeley National Laboratory, Berkeley, California; ⁴Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, California; and ⁵Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov>

Project Goals: See goals for abstract 36.

Microarrays fabricated with the genes encoding key, functional enzymes involved in various biological and geochemical cycling processes are referred to as functional gene arrays (FGAs). On the basis of GeoChip 2.0, which contains 24,243 oligonucleotide (50mer)

probes and covers > 10,000 gene sequences in >150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation, a new generation of GeoChip (GeoChip 3.0) has been developed. GeoChip 3.0 has several new features compared to GeoChip 2.0. First, GeoChip 3.0 is expected to cover >37,000 gene sequences for more than 300 gene families, and such a coverage allows us to obtain more information about microbial communities and analyze more diverse environmental samples. Second, the homology of automatically retrieved sequences by key words is verified by HUMMER using seed sequences so that unrelated sequences are removed. Third, GeoChip 3.0 includes phylogenetic markers, such as *gyrB*. Fourth, a software package (including databases) has been developed for sequence retrieval, probe and array design, probe verification, array construction, array data analysis, information storage, and automatic update, which greatly facilitate the management of such complicated array, especially for future update. Fifth, a universal standard has been implemented in GeoChip 3.0 so that data normalization and comparison can be conducted. Sixth, a genomic standard is also used to quantitatively analyze gene abundance. Finally, GeoChip 3.0 also includes GeoChip 2.0 probes, and those GeoChip 2.0 probes are checked against new databases. Disqualified probes are flagged. GeoChip 3.0 will provide more capability for studying biogeochemical processes and functional activities of microbial communities important to human health, agriculture, energy, global climate change, ecosystem management, and environmental cleanup and restoration. It is also particularly useful for providing direct linkages of microbial genes/populations to ecosystem processes and functions. In the new proposal, we will explore to discover the molecular mechanisms by which microbial community structure, function, and stability affects stress response and activity in the lab and in the field. GeoChip 3.0 or similar but more specific functional gene arrays will directly examine the natural community dynamics in the field for correlation to laboratory observations (Aim 3 of the new proposal), and will facilitate our understanding of the molecular basis of microbial community formation, stability, resilience, specificity, activity and ecology. In addition, all GeoChip data can be integrated to trace the path from genome to community phenotypes.

* Presenting author

53

GTL

Comparative Metagenomics of Microbial Communities from Pristine and Contaminated Groundwater

Christopher L. Hemme,^{1,2*} Ye Deng,¹ Terry J. Gentry,³ Liyou Wu,^{1,2} **Matthew W. Fields**,^{2,4} Susannah Green-Tringe,⁵ Chris Detter,⁵ Kerrie Barry,⁵ Nikos Kyrpides,⁵ David Watson,⁶ Paul Richardson,⁵ **Terry C. Hazen**,^{2,7} James Tiedje,⁸ Eddy Rubin,⁵ and **Jizhong Zhou**^{1,2} (jzhou@ou.edu)

¹Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma; ²Virtual Institute of Microbial Stress and Survival <http://vimss.lbl.gov/>; ³Department Soil and Crop Sciences, Texas A&M University, College Station, Texas; ⁴Department of Microbiology, Montana State University, Bozeman, Montana; ⁵DOE Joint Genome Institute, Walnut Creek, California; ⁶Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; ⁷Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, California; and ⁸Center for Microbial Ecology, Michigan State University, East Lansing, Michigan

Project Goals: See goals for abstract 36.

To better understand the evolutionary and physiological mechanisms by which microbial communities adapt to environmental stress, community-scale genomic analyses are necessary. To this end, microbial community DNA isolated from contaminated groundwater located at the U.S. Dept. of Energy Field Research Center (FRC) (Area 3, Well FW106) in Oak Ridge, TN, was analyzed to determine the effects of chronic exposure to multiple stressors (including low pH, nitrate, uranium and organic solvents) on the nascent microbial community structure. Consistent with previous 16S rRNA gene surveys conducted at the FRC, metagenomic analysis indicates a very low diversity community (~13 OTU) dominated by clonal denitrifying γ - and β -proteobacterial populations. Metabolic reconstruction of the dominant γ -proteobacterial species reveals adaptations for specific geochemical parameters including the following: denitrification and ammonium assimilation pathways; mechanisms for resistance to low pH; pathways for degradation of organic compounds such as 1,2-dichloroethene, acetone, butanol, methanol and formaldehyde; accumulation of multiple heavy metal efflux systems (*czcABC*, *czcD*, *cadA*-family, *mer* operon genes, etc.). Analysis indicates that lateral gene transfer is the predominant mechanism of introducing genetic variation into the community,

resulting in the lateral acquisition of geochemical resistance genes (e.g. acetone carboxylation, heavy metal efflux systems etc.). A particularly interesting adaptation is the number and variety of toxic efflux genes identified in the FW106 metagenome. In particular, genes encoding NarK nitrate/nitrite antiporters, CzcABC divalent cation transporters and CzcD divalent cation transporters are highly abundant, suggesting that these genes are important in community stress response. The most likely physiological effect of this accumulation of transporters would be to increase the baseline rate of toxic efflux from the cell. Rapid efflux of toxins in conjunction with specific degradation and detoxification pathways, thus appears to be a major survival strategy under stressed conditions.

The FW106 sample was compared to a second groundwater metagenome from a pristine FRC site (FW301) to determine differences between the two communities. In contrast to the low species diversity of FW106, the FW301 is represented by multiple phyla including all 5 classes of proteobacteria, *Planctomycetes*, *Chloriflexi*, *Actinobacteria*, *Acidobacteria*, *Bacteroidetes* and *Firmicutes*. In contrast to the FW106 sample which resulted in significant read assembly, the FW301 sample is composed largely of single reads that do not assemble into contigs (95%). Interestingly, most of the geochemical resistance mechanisms identified in FW106 are also present in FW301 as well as other resistance mechanisms (e.g. aromatic degradation) that are not found in FW106. Abundance profiling of geochemical and cytochrome genes between FW106 and FW301 and between FW106 and the acid mine drainage (AMD) metagenome show distinct environmental signatures between the samples, including the accumulation of cytochrome *c553* genes in the FW106 metagenome which may play a role in metal resistance. The AMD metagenome also showed an abundance of CzcABC genes, suggesting that accumulation of these genes may be an important resistance mechanism in metal-stressed communities. The identification of FW106 stress response genes will also permit more in-depth studies of the dynamics of stressed FRC communities using functional genomics tools such as the GeoChip.

* Presenting author

54

Comparative Genomics of Ethanol-Producing *Thermoanaerobacter* Species

Christopher L. Hemme,^{1,7*} **Matthew W. Fields,**^{2,7} Qiang He,³ Zhiguo Fang,¹ Zhili He,^{1,7} J.C. Detter,⁴ Kerrie Barry,⁴ Miriam Land,⁴ Loren Hauser,⁴ Alla Lapidus,⁵ Cliff S. Han,⁶ Paul Richardson,⁵ Eddy Rubin,⁶ and **Jizhong Zhou**^{1,7} (jzhou@ou.edu)

¹Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma; ²Department of Microbiology, Montana State University, Bozeman, Montana; ³Department of Civil and Environmental Engineering, Temple University, Philadelphia, Pennsylvania; ⁴Genome Analysis and Systems Modeling Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; ⁵DOE Joint Genome Institute, Walnut Creek, California; ⁶Los Alamos National Laboratory, Los Alamos, New Mexico; and ⁷VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>

Project Goals: See goals for abstract 36.

Recent global fluctuations in the supply and demand of petroleum have prompted a call for increased research into biologically-derived fuels (biofuels). Of particular interest are those processes involving production of biofuels from cellulosic biomass. The conversion of biomass to biofuels is critically dependent on the flux of carbon through the fermentation pathways. To better understand the carbon pathways involved in this process, the genomes of two strains of ethanol-producing *Thermoanaerobacter* species (*T. pseudoethanolicus* 39E and *T. ethanolicus* X514) have been sequenced and compared to a third species, *Caldanaerobacter subterraneus* supsp. *tengcongensis* (formerly *Thermoanaerobacter tengcongensis*). Strain 39E is a well-characterized strain isolated from a Yellowstone hot spring and is capable of degrading xylan and starch to ethanol at high yields. Strain X514 is a metal-reducing bacteria isolated from the deep subsurface and is predicted to have been geographically isolated from 39E for ~250 MY. Both strains show a high degree of gene conservation but marked genomic rearrangements compared to each other and to *C. subterraneus*. The common *Thermoanaerobacter* ancestor underwent a genome rearrangement around the origin of replication following divergence from *C. subterraneus*. Additional large-scale rearrangements occurred in X514 following its divergence from 39E which has resulted in a strong

GTL

strand asymmetry in X514, with 62.3% of the X514 genes being located on the leading strand. A survey of unique genes between the strains reveals lineage-specific gene expansions in the two strains including individual unique sugar transporter profiles and an increased number of P-type metal translocating ATPase genes in X514. Several of these gene clusters are associated with phage genes, suggesting phage-mediated lateral gene transfer. The activity of phage in the X514 environment is also evidenced by a significantly expanded CRISPR region in X514 compared to 39E, which implies phage resistance activities. X514 also encodes a complete Vitamin B₁₂ biosynthesis pathway that is partially lacking in 39E and *C. subterraneus*. This pathway is associated with a large genomic rearrangement observed between the two *Thermoanaerobacter* strains. Exogenous B₁₂ has been shown to increase ethanol yields in certain strains of *Clostridium thermocellum* and thus the ability to synthesize B₁₂ *de novo* may contribute to the increased ethanol yields observed with X514.

Comparisons of ethanol yields of 39E and X514 in both mono- and batch culture suggest that X514 is a more efficient ethanol producer than 39E. Because X514 was isolated from a presumably nutrient poor environment, it is hypothesized that the strain has evolved high-affinity carbon scavenging mechanisms that may contribute to the observed high ethanol yields. Such mechanisms may manifest as novel genes or in modifications to existing genes. Metabolic reconstruction of the *Thermoanaerobacter* species reveals insights into carbon metabolism and niche adaptation of the two strains. Both strains are capable of metabolizing glucose and xylan to ethanol with a novel bifunctional secondary alcohol dehydrogenase serving as the terminal enzyme in the pathway. Differences are noted in the carbon metabolism pathways of the two strains, including a complete KDPG metabolism pathway in 39E and the lack of a complete methylglyoxal shunt in X514. Abundance profiles of leading vs lagging strand genes based on COG categories for both 39E and X514 show an overabundance of carbon transport and metabolism genes on the leading strand of X514 compared to the lagging strand of X514 and the leading strand of 39E, suggesting that carbon metabolism genes may be more highly expressed in X514. However, an initial analysis of codon adaptation indices (CAI), a possible measure of gene expression, does not support this hypothesis. CAI analysis, however, does show that cellobiose-specific PTS component genes are highly expressed in both strains which may contribute to the increased ethanol production of these strains in coculture with cellulosic bacteria.

* Presenting author

To complement this research, a request for sequencing the genomes of an additional 20 ethanol-producing *Clostridia* strains has been approved by JGI. Strains were chosen from among the genera *Clostridium*, *Thermoanaerobacter*, *Thermoanaerobacterium* and *Acetivibrio* based on prior knowledge, phylogeny, unique physiology and industrial applications. The expansion of the genomic database of industrially-important *Clostridia* is expected to provide substantial benefits in the understanding of this class of organisms. Additional computational and functional genomic analyses are currently being conducted to further elucidate the mechanisms of carbon usage and ethanol production in these bacteria and to correlate the computational predictions with experimental gene expression data. A better understanding of the expression of biomass conversion genes will aid in future efforts to engineer these bacteria for industrial use.

55

Metabolic Flux Analysis of *Shewanella* spp Central Carbon Metabolism Reveals Evolutionary, Genetic, and Environmental Robustness

Yinjie J. Tang,^{1,2*} Hector Garcia Martin,³ Paramvir S. Dehal,^{1,2} Adam Deutschbauer,^{1,2} Xavier Llorca,⁴ Adam Meadows,⁵ Adam P. Arkin^{1,2,6} (aparkin@lbl.gov), and Jay D. Keasling^{1,2,6,7}

¹Virtual Institute for Microbial Stress and Survival <http://vimss.lbl.gov>, Lawrence Berkeley National Laboratory, Berkeley, California; ²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; ³Microbial Ecology Program, DOE Joint Genome Institute, Walnut Creek, California; ⁴National Center for Supercomputing Applications, University of Illinois, Urbana-Champaign, Illinois; ⁵Genentech, Inc., South San Francisco, California; ⁶Department of Bioengineering, University of California, Berkeley, California; and ⁷Department of Chemical Engineering, University of California, Berkeley, California

Project Goals: See goals for abstract 36.

A major goal of the Environmental Stress Pathway Project (ESPP) is to uncover the mechanisms that allow microbes to sustain activity under the environmental stress seen in field settings. In this study, we used ¹³C isotopomer analysis to elucidate the metabolic flux through central metabolism of eight *Shewanella* species. Metabolite fluxes link genes, proteins and metabolites

to macroscopic biological functions. In spite of its importance, only a few, not thoroughly tested, general principles have been proposed to predict and understand the flux configuration of an organism. Among those general principles, robustness of central metabolism has been reported with respect to genetic perturbation. Here we show that the relative metabolic flux distributions are very similar for phylogenetically and environmentally diverse members of the *Shewanella* genus. This phylogenetic robustness suggests understanding microbial fluxomics in terms of metabolic types (or metatypes), as opposed to phylotypes. In addition to phylogenetic, environmental and genetic robustness our data shows flexibility in the relative flux profiles when adapting to different carbon sources.

56

VIMSS Computational Core

Paramvir S. Dehal^{1,2*} (PSDehal@lbl.gov), Eric J. Alm,^{1,3} Dylan Chivian,^{1,2} Katherine H. Huang,^{1,2} Y. Wayne Huang,^{1,2} Marcin P. Joachimiak,^{1,2} Keith Keller,^{1,4} Morgan N. Price,^{1,2} and Adam P. Arkin^{1,2,4,5,6}

¹Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov/>; ²Lawrence Berkeley National Laboratory, Berkeley, California; ³Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts; ⁴University of California, Berkeley, California; ⁵Howard Hughes Medical Institute, Chevy Chase, Maryland; and ⁶Department of Bioengineering, University of California, Berkeley, California

Project Goals: See goals for abstract 36.

Background: The VIMSS Computational Core group is responsible for data management, data integration, data analysis, and comparative and evolutionary genomic analysis of the data for the VIMSS project. We have expanded and extended our existing tools sets for comparative and evolutionary genomics and microarray analysis as well as creating new tools for our proteomic and metabolomic data sets. Our analysis has been incorporated into our comparative genomics website MicrobesOnline (<http://www.microbesonline.org>) and made available to the wider research community. By taking advantage of the diverse functional and comparative datasets, we have been able to pursue large evolutionary studies.

Data Analysis: During the course of analysis of various stress responses of DvH, the computational core has

* Presenting author

continued to develop new statistical analyses of data that take advantage of the predicted regulatory structures (operons, regulons, etc.) from our comparative analyses. This year we have used these analyses to investigate the response of DvH to oxygen stress and pH stress. Our analysis has focused on the combined results from both transcriptomic and proteomic datasets to interpret oxygen stress. Additionally, we have worked with metabolomic datasets within the framework of predicted metabolic activities to find missing pathway members.

Data Management: All data generated by ESPP continues to be stored in our Experimental Information and Data Repository (<http://vimss.lbl.gov/EIDR/>). Researchers have access to datasets from biomass production, growth curves, image data, mass spec data, phenotype microarray data and transcriptomic, proteomic and metabolomic data. New functionality has been added for storage of information relating to mutants and protein complex data, in addition to new visualization for assessing existing data sets such as the phenotype microarrays.

The MicrobesOnline Database: The MicrobesOnline database (<http://www.microbesonline.org>) currently holds over 700 microbial genomes and will be updated quarterly, providing an important comparative genomics resource to the community. New functionality added this year includes the addition of a thousands of phage genomes and plasmids, an updated user interface for the phylogenetic tree based genome browser that allows users to view their genes and genomes of interest within an evolutionary framework, tools to compare multiple microarray expression data across genes and genomes, addition of external microarray data from the Many Microbial Microarrays Database, integration with the RegTransBase of experimentally verified regulatory binding sites and links to three dimensional protein structures of proteins and their close relatives.

MicrobesOnline continues to provide an interface for genome annotation, which like all the tools reported here, is freely available to the scientific community. To keep up with the rapidly expanding set of sequenced genomes, we have begun to investigate methods for accelerating our annotation pipeline. In particular we have completed work on methods to speed up the most time consuming process, homology searching through HMM alignments and all against all BLAST. These methods now enable us to deal with the many millions of gene sequences generated from metagenomics.

Over the next year, several new features will be added to the MicrobesOnline resource. Microarray expression data will be added from the NCBI GEO database, in addition

to datasets generated from the VIMSS team. To supplement the analysis tools we already have, enrichment of functional genes and operon-wise analysis, we will provide tools for comparing multiple experiments across multiple genomes. We will also expand our regulatory binding motif search to incorporate co-expression data to support predictions.

Evolutionary Analysis: The computational core continues work on understanding the evolution of regulatory networks. Transcription factors form large paralogous families and have complex evolutionary histories. Our analysis shows that putative orthology derived from bidirectional best hits across distantly related bacteria are usually not true evolutionary orthologs. Additionally, these false orthologs usually respond to different signals and regulate distinct pathways. Even in more closely related genomes, such as *E. coli* and *Shewanella oneidensis*, bidirectional best hits have a high error rate. By studying transcription factors with phylogenetic trees, we show that through the use of gene-regulon correlations, together with sequence analysis of promoter regions for confirmation, bacterial regulatory networks may evolve more rapidly than previously thought.

57

GTL

MicrobesOnline: An Integrated Portal for Comparative Microbial Functional Genomics

Marcin P. Joachimiak,^{1,2*} Y. Wayne Huang,^{1,2} Katherine H. Huang,^{1,2} **Eric J. Alm**,^{1,3} Dylan Chivian,^{1,2} Paramvir S. Dehal,^{1,2} Keith Keller,^{1,4} Morgan N. Price,^{1,2} and **Adam P. Arkin**^{1,2,4,5} (aparkin@lbl.gov)

¹Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov/>; ²Lawrence Berkeley National Laboratory, Berkeley, California; ³Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts; ⁴University of California, Berkeley, California; and ⁵Department of Bioengineering, University of California, Berkeley, California

Project Goals: See goals for abstract 36.

The Virtual Institute for Microbial Stress and Survival (VIMSS, <http://vimss.lbl.gov/>) funded by the Dept. of Energy's Genomics:GTL Program, is dedicated to using integrated environmental, functional genomic, and comparative sequence and phylogeny data to understand mechanisms by which microbes and microbial communi-

* Presenting author

ties survive in uncertain environments while carrying out processes of interest for bioremediation and energy generation. To support this work, VIMSS has developed a Web portal, along with computational analysis and an underlying database, for comparative functional genomics of bacteria and archaea. MicrobesOnline (<http://www.microbesonline.org>) has been enabling comparative genome analysis since 2003. The portal currently includes 702 complete microbial genomes (662 bacterial, 40 archaeal, 1770 viruses, 268 plasmids) and offers a suite of analysis and tools including: comprehensive gene family and domain annotations, information on three dimensional structure representatives and similarities, an interactive gene expression heatmap browser along with functionality for gene expression profile searches, a multi-species genome browser, operon and regulon predictions, a combined gene and species phylogeny browser, a gene ontology browser, a metabolic pathway browser, a workbench for sequence analysis (including sequence motif detection, motif searches, sequence alignment and phylogeny reconstruction), and capabilities for community annotation of genomes.

VIMSS integrates functional genomic data and provides novel web-based viewing and analysis tools for gene expression microarray, proteomic, metabolomic, and phenotype microarray data. Currently, these data are mostly project generated for wild-type and mutants of *Desulfovibrio vulgaris* and *Shewanella oneidensis* exposed to stress conditions found at DOE field sites. However, the organism scope is being expanded, and recently the *E. coli* gene expression data compendium from the Many Microbes Microarray Database (M3D) has been imported into MicrobesOnline. Additionally VIMSS has developed capabilities to analyze microarray experiments performed on multiple species simultaneously. Selecting an organism or gene of interest in MicrobesOnline leads to information about and data viewers for VIMSS experiments conducted on that organism and involving that gene or gene product. It is possible to view microarray data from multiple stress conditions as an interactive heatmap and to analyze correlations between gene expression results from different experiments. Among the major new features is the ability to search the microarray data compendium for genes with gene expression profiles similar to a query expression profile (either based on a gene or set of genes). Such new compendium-wide functionalities allow to observe patterns in gene expression changes across multiple conditions and genes, and to search for similarities to these patterns. The information integration and analysis performed by VIMSS serves not only to generate insights into the stress responses and their regulation in these microorganisms, but also to document VIMSS experiments, allow contextual

access to experimental data, and facilitate the planning of future experiments. VIMSS also is incorporating into MicrobesOnline publicly available functional genomics data from published research, so as to centralize and synergize data on and analysis of microbial physiology and ecology in a unified comparative functional genomic framework.

58 GTL The Analysis and Expansion of Regulatory Binding Site Data in a Wide Range of Bacteria Using a Semi- Automatic System—RegTransBase

Michael J. Cipriano,¹ Alexei E. Kazakov,² Dmitry Ravcheev,² **Adam P. Arkin**^{1,3,4,5} (aparkin@lbl.gov), Mikhail S. Gelfand,^{2,6} and Inna Dubchak^{1,5,7*}

¹Lawrence Berkeley National Laboratory, Berkeley, California; ²Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia; ³Howard Hughes Medical Institute, Chevy Chase, Maryland; ⁴Department of Bioengineering, University of California, Berkeley, California; ⁵Virtual Institute of Microbial Stress and Survival <http://vimss.lbl.gov/>; ⁶Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia; and ⁷DOE Joint Genome Institute, Walnut Creek, California

Project Goals: See goals for abstract 36.

RegTransBase, a database describing regulatory interactions in prokaryotes, has been developed as a component of the MicrobesOnline/RegTransBase framework successfully used for interpretation of microbial stress response and metal reduction pathways. It is manually curated and based on published scientific literature. RegTransBase describes a large number of regulatory interactions and contains experimental data which investigates regulation with known elements. It is available at <http://regtransbase.lbl.gov>. Currently, the database content is derived from more than 4000 relevant articles describing over 9000 experiments in relation to 155 microbes. It contains data on the regulation of ~14000 genes and evidence for ~7500 interactions with ~850 regulators.

RegTransBase additionally provides an expertly curated library of alignments of known transcription factor binding sites covering a wide range of bacterial species. Each alignment contains information as to the transcription factor which binds the DNA sequence, the exact location

* Presenting author

of the binding site on a published genome, and links to published articles.

RegTransBase builds upon these alignments by containing a set of computational modules for the comparative analysis of regulons among related organisms. These modules guide a user through the appropriate steps of transferring known or high confidence regulatory binding site results to other microbial organisms, allowing them to study many organisms at one time, while warning of analysis possibly producing low confidence results, and providing them with sound default parameters.

There is an increasingly tight coupling of RegTransBase with MicrobesOnline in reporting cis-regulatory sites and regulatory interactions, and integrating RegTransBase searches into MicrobesOnLine cart functions.

most annotations of bacterial regulators are probably incorrect.

When we analyzed the histories of regulatory interactions, we found that the evolution of regulation by duplication was rare, and surprisingly, many of the regulatory interactions that are shared between paralogs result from convergent evolution. Furthermore, horizontally transferred genes are more likely than other genes to be regulated by multiple regulators, and most of this complex regulation probably evolved after the transfer. Finally, gene regulation is often not conserved, even within the gamma-Proteobacteria. Our results suggest that the bacterial regulatory network is evolving rapidly under positive selection. Such rapid rewiring of gene regulation may be crucial for adaptation to new niches.

59 Evolutionary History of Gene Regulation in Bacteria

GTL

Morgan N. Price,^{1,2} Paramvir S. Dehal,^{1,2*} and Adam P. Arkin^{1,2,3} (aparkin@lbl.gov)

¹Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov/>; ²Lawrence Berkeley National Laboratory, Berkeley, California; and ³Department of Bioengineering, University of California, Berkeley, California

Project Goals: See goals for abstract 36.

Analysis of gene regulation in ESPP bacteria relies on comparisons to model organisms, and hence on assumptions about how gene regulation evolves. To test these assumptions we examined the evolutionary histories of transcription factors and of regulatory interactions from the model bacterium *Escherichia coli* K12. We show that although most transcription factors have paralogs, these usually arose by horizontal gene transfer rather than by duplication within the *E. coli* lineage, as previously believed. Most neighbor regulators — regulators that are adjacent to genes that they regulate — were acquired by horizontal gene transfer, while most global regulators evolved vertically within the gamma-Proteobacteria. Neighbor regulators are often acquired together with the operon that they regulate, which suggests that the proximity is maintained by repeated transfers, and also aids the prediction of the regulators' function. Because of the complex evolutionary histories of most transcription factors, bidirectional best hits tend to be misleading, and

60 MicroCOSM: Phylogenetic Classification of Metagenomic Data Using Microbial Clade-Oriented Sequence Markers

GTL

Dylan Chivian^{1,2*} (DCChivian@lbl.gov), Paramvir S. Dehal,^{1,2} and Adam P. Arkin^{1,2,3}

¹Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov/>; ²Lawrence Berkeley National Laboratory, Berkeley, California; and ³University of California, Berkeley, California

Project Goals: See goals for abstract 36.

The VIMSS/ESPP2 project requires understanding of the microbial communities at contaminated field sites and, among other methods, will employ metagenomics in this endeavor. Metagenomics projects that seek to elucidate the population structure of microbial ecosystems are faced with the related computational challenges of classifying the sequences obtained and quantifying which organisms are present within a sample. Individually low-proportion species usually make up a large fraction of microbial communities, complicating their classification and quantification using traditional phylogenetic marker approaches. Such species usually don't yield sufficient read depth to assemble into longer sequences, leaving fragments that rarely contain traditional markers such as the small subunit (SSU) rRNA gene. BLAST-based approaches for analysis of metagenomic sequences [1] compensate for this rarity of traditional markers, but may be confounded by genes that are subject to horizontal transfer or duplication. Another approach instead makes

* Presenting author

use only of reliable non-transferred single-copy genes [2] to classify and quantify the organisms present within a sample, but the application has so far been limited to the use of a fairly small set of universal genes found in all organisms. In this work, we have extended the latter approach, boosting the set of reliable marker genes from only about 30–40 universal genes to several hundred by identifying sets of single-copy genes that are not subject to inter-clade horizontal transfer through investigation of finished bacterial and archaeal genomes. These clade-oriented sequence markers allow for a method, which we have named “MicroCOSM”, that greatly increases the probability that a marker will be found in any given sequence and therefore offers improved coverage for phylogenetic classification and quantification of microbial types in an environmental sample.

References

1. Huson D.H., Auch A.F., Qi J., Schuster S.C. (2007) “MEGAN analysis of metagenomic data.” *Genome Res.* **17**(3):377–86.
2. von Mering C., Hugenholtz P., Raes J., Tringe S.G., Doerks T., Jensen L.J., Ward N., Bork P. (2007) “Quantitative phylogenetic assessment of microbial communities in diverse environments.” *Science* **315**(5815):1126–30.

61

AdaptML: A Maximum Likelihood Framework for Integrating Evolutionary and Ecological Reconstructions

Lawrence David,¹ Dana Hunt,² Dirk Gevers,^{2,3,4} Sarah Preheim,² Chiachi Hwang,^{5,7} **Matthew Fields**,^{5,7} Martin Polz,² and **Eric Alm**^{2,6,7*} (ejalm@mit.edu)

¹Computational and Systems Biology Initiative, ²Department of Civil and Environmental Engineering, ³Laboratory of Microbiology and ⁴Bioinformatics and Evolutionary Genomics Group, Ghent University, Ghent, Belgium; ⁵Department of Microbiology, Montana State University, Bozeman, Montana; ⁶Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts; and ⁷Virtual Institute of Microbial Stress and Survival (<http://vimss.lbl.gov>)

Project Goals: See goals for abstract 36.

To date, modeling the behavior of cellular networks under laboratory conditions has received more attention

than modeling how ecological factors affect diversity in natural environments. As we move toward the ultimate goal of integrating laboratory model organism studies with field data, a key challenge will be identifying the geochemical/ecological factors that underlie community diversity, and the phylogenetic boundaries of natural ecological populations. Thus, computational frameworks for automatically learning models of sequence evolution in the context of metadata (*e.g.*, site geochemistry/ecology) will need to be developed. We present here one such framework: *AdaptML*, a maximum-likelihood-based tool for studying both the sequence evolution and ecological history of a set of gene sequences. To perform this latter task, *AdaptML* employs a hidden-Markov-model-like strategy of assigning gene sequences to unseen states we term “habitats.” These habitats are inferred automatically and designed to recapitulate sequence partitioning observed in the wild. *AdaptML* was initially developed and tested using data from 1027 strains of marine *Vibrio hsp60* gene sequences harvested off the coast of Maine. We show here how *AdaptML* can be used to analyze this dataset and to help build models of *Vibrio* resource partitioning. We have recently applied *AdaptML* to a 16S library collected from a DOE FRC site as part of the ESPP2 project with pre and post biostimulation time points, as a first step toward identifying ecological factors that drive changes in taxonomic diversity and thus organismal fitness in an environment relevant to DOE missions.

GTL

62

VIMSS:ESPP2 Scientific Research Project Management

Amy Shutkin (ashutkin@lbl.gov),^{1,2*} **Adam P. Arkin**,^{1,2,3} and **Terry C. Hazen**^{1,2}

¹Lawrence Berkeley National Laboratory, Berkeley, California; ²VIMSS (Virtual Institute of Microbial Stress and Survival) <http://vimss.lbl.gov/>; and ³University of California, Berkeley, California

Project Goals: See goals for abstract 36.

The success of a mature scientific research program depends on maintaining an agile Performance Monitoring Project Management structure, which served the flagship VIMSS:ESPP effectively, while implementing more formal risk management oversight practices as proposed new technologies are integrated into the pre-existing core research groups. Project Management offers the stability of a rational and logical process for managing work in a virtual research institute. This (re)new

GTL

* Presenting author

research project initiated at three national laboratories and seven universities from coast to coast and is comprised of ~75 individuals collaborating within three core research groups: Applied Environmental Microbiology Core (AEMC), Functional Genomics and Imaging Core (FGIC) and Computational and Systems Biology Core (CSBC). We have found that there is a balance between tight project management to create synergy, focus and continuity to the project and well-tracked individual-investigator-driven initiatives and follow-up that must be struck to maintain creative engagement and productivity of the project team, all while remaining vigilant against scope creep. It has been our experience that frequent and rapid communication at different levels with different media is critical to exploiting the scale and diversity of the team project's capabilities for this multidisciplinary, multi-institutional collaboration.

Included topics in this overview are: Project Management Plan Development and Implementation, Performance Monitoring and Communications, Schedule Development and Execution, Cost Estimation, and Risk Identification and Management.

63 Systematic Identification of Regulatory Mapping and Optimal Metabolic Engineering Strategies in *Shewanella oneidensis* MR-1

GTL

D.K. Byrne^{1*} (dbyrne@bu.edu), Q.K. Beg^{1*} (qasimbeg@bu.edu), M.E. Driscoll,¹ F.S. Juhn,¹ Y. Shen², J.J. Faith,¹ I. Paschalidis,² **D. Segre,¹** and **T.S. Gardner¹**

¹Department of Biomedical Engineering and

²Department of Systems Engineering, Boston University, Boston, Massachusetts

Project Goals: The project goals are to: (1) construct a genome-scale model of transcriptional regulatory pathways controlling metabolic enzymes and stress response systems, (2) complete and validate a metabolic flux model of *Shewanella oneidensis*, and (3) apply a constraints-based analysis to integrated metabolic and regulatory models to optimize electron flux.

Shewanella oneidensis MR-1 is an environmentally ubiquitous, metabolically versatile microbe with a broad capacity for the reduction of metals. These desirable traits have made *Shewanella* a leading candidate for use in microbial fuel cells (MFCs) and environmental remedia-

tion of toxic waste. However, practical engineering implementations are limited by our ability to systematically characterize and control key regulatory and metabolic aspects of the organism's physiology.

To this end, we have designed an Affymetrix microarray for *Shewanella* and profiled its genome-wide expression across 255 conditions [1] by varying carbon sources, metallic species, and physiologically relevant factors. We applied our CLR algorithm [2] on this expression compendium to infer the first full-scale regulatory network for *Shewanella*. The resultant network revealed several previously uncharacterized genes as central regulators of the respiratory machinery. *Shewanella* strains containing knock-outs of these genes are currently being investigated to confirm our predictions.

Also, we have developed computational methods to identify optimal strategies to improve *Shewanella* phenotypes related to electrical current production in MFCs. Using a genome-scale metabolic model, developed by the *Shewanella* Federation and led by Jennifer Reed (University of Wisconsin-Madison), we developed and applied algorithms based on linear and nonlinear programming theory to predict optimal nutrients, nutrient compositions, and gene knock-outs that maximize the electrical power output and coulombic efficiencies that *Shewanella* strains produce. Validation of these predictions using batch and continuous bioreactor experiments on wild-type and mutant *Shewanella* strains are currently under way.

The complementary regulatory and metabolic methods developed here will extend current knowledge of the genes, regulators, and metabolites that underlie *Shewanella's* unique respiratory system, and the environmental signals to which they respond. Furthermore, our predictions provide strategies to optimize *Shewanella* for improved metal reduction in real-world applications.

References

1. JJ Faith, ME Driscoll, VA Fusaro, EJ Cosgrove, B Hayete, FS Juhn, SJ Schneider, TS Gardner. (2007) Many Microbe Microarrays Database: uniformly normalized Affymetrix compendia with structured experimental metadata. *Nucleic Acids Research*. [Epub ahead of print].
2. JJ Faith, B Hayete, JT Thaden, I Mogno, J Wierzbowski, G Cottarel, S Kasif, JJ Collins, TS Gardner. (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulatory interactions from a compendium of expression profiles. *PLoS Biology*. 5: e8.

* Presenting author

Towards Genomic Encyclopedia of Carbohydrate Utilization: Case Study in *Shewanella*

Dmitry Rodionov,¹ Chen Yang,¹ Xiaoqing Li,¹ Andrei Osterman^{1,2*} (osterman@burnham.org), Olga Zagnitko,² Ross Overbeek,² Samantha Reed,³ Margaret Romine,³ Yanbing Wang,⁴ Anna Y. Obratsova,⁴ and **Kenneth H. Nealson**⁴ (knealson@usc.edu)

¹Burnham Institute for Medical Research, La Jolla, California; ²Fellowship for Interpretation of Genomes, Burr Ridge, Illinois; ³Pacific Northwest National Laboratory, Richland, Washington; and ⁴Department of Earth Sciences, University of Southern California, Los Angeles, California

Project Goals: This project is a component of the *Shewanella* Federation studies aimed at better understanding of the ecophysiology and speciation of this important genus. The main goal of the presented study is to combine genomic reconstruction with physiological experiments to establish and characterize pathways involved in utilization of carbon and energy sources in a collection of *Shewanella* species with completely sequenced genomes.

Carbohydrates are a primary source of carbon and energy for many divergent bacteria. Tremendous variations in carbohydrate pools in different ecological reservoirs are matched by variations within the carbohydrate utilization machinery observed even between closely related species. Our long-range goal is to establish a capability of accurate and comprehensive reconstruction of this machinery from genomic data for hundreds (soon to be thousands) sequenced bacterial genomes. This capability would strongly impact fundamental understanding of microbial ecophysiology and foster applications in bioengineering. Despite the substantial knowledge accumulated in this field, our current ability to project the respective genes and pathways from a handful of model species to others is hampered by the abundance of paralogous protein families of varied specificity, nonorthologous gene replacements and pathway variations. This leads to the accumulation and propagation of imprecise and, often, incorrect genomic annotations in public archives. We address this challenge by combining a subsystems-based approach to genome annotation with experimental validation of selected bioinformatic predictions. The subsystem analysis (as implemented in The SEED genomic platform, <http://theseed.uchicago.edu/>)

is a highly parallel genomic reconstruction of metabolic pathways in hundreds of divergent microbial species supported by the extensive exploration of genome context (conserved operons and regulons). In addition to improving the quality of annotations, this approach allows us to make conjectures about previously uncharacterized genes, pathways and phenotypes. The key inferences are tested by genetic, biochemical and physiological experiments in model species.

We applied this integrated approach to systematically map mono- and disaccharide catabolic pathways in 12 species of *Shewanella* with completely sequenced genomes. This project is a component of the *Shewanella* Federation studies aimed at better understanding of the ecophysiology and speciation of this important genus. We used a set of nearly 250 protein families including transporters, transcriptional regulators and several classes of metabolic enzymes (e.g. sugar kinases, oxidoreductases, epimerases, aldolases, etc) from a collection of carbohydrate utilization subsystems in The SEED to scan for homologs in *Shewanella* genomes. Identified candidate genes were subject of further genome context analysis for their accurate functional assignment and reconstruction of respective pathways. This analysis detected substantial variations in a *sugar diet* between different *Shewanella* species reflecting various aspects of their ecophysiology and evolutionary history. However, more striking are the differences revealed by comparison with a classical model system of *E. coli*. These differences are manifested at various levels, from the presence or absence of certain sugar catabolic pathways to a dramatically different organization of transcription regulatory networks in the central carbon metabolism. The results of this analysis included prediction of several novel variants of carbohydrate utilization pathways (eg for N-acetylglucosamine, sucrose, cellobiose, arabinose and glycerate) and tentative functional assignments for previously uncharacterized gene families (eg within GlcNAc operon SO3503-3507 in *S. oneidensis*, sucrose operon Sfri3988-3991 in *S. firgidi-marina* and cellobiose operon Sbal0541-0545 in *S. baltica*). These predictions were verified by phenotype analysis, genetic complementation and biochemical characterization of purified recombinant enzymes. In addition to the specific knowledge of carbohydrate catabolism in the *Shewanella* genus, this study led to a substantial expansion of our current version of the Genomic Encyclopedia of Carbohydrate Utilization. A systematic iterative application of this approach to multiple taxonomic groups of bacteria will further enhance this knowledge base providing an adequate support for efficient analysis of newly sequenced genomes as well as of the emerging metagenomic data.

* Presenting author

65

Discovery of Novel Machinery for Lactate Utilization by *Shewanella oneidensis* MR-1

Grigoriy Pinchuk,¹ Dmitry Rodionov^{2*} (rodionov@burnham.org), Chen Yang,² Xiaoqing Li,² Samantha Reed,¹ Margaret Romine,¹ Alex Beliaev,¹ and **Andrei Osterman**^{2,3} (osterman@burnham.org)

¹Pacific Northwest National Laboratory, Richland, Washington; ²Burnham Institute for Medical Research, La Jolla, California; and ³Fellowship for Interpretation of Genomes, Burr Ridge, Illinois

Project Goals: This project is a component of the *Shewanella* Federation and, as such, contributes to the overall goal of applying the genomic tools to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.

Lactate is one of the major fermentative metabolism products for many microorganisms and is the most frequently used substrate for experimental studies of respiratory metabolism in *Shewanella oneidensis* MR-1. Consequently, the metabolism of lactate is a key component of the systems-level conceptual model (under development by the *Shewanella* Federation) linking electron transfer networks and central/peripheral carbon metabolism pathways of MR-1. Whereas physiological data has demonstrated a robust growth of *S. oneidensis* on both D- and L- forms of lactate, its genome does not contain orthologs of classical lactate dehydrogenases (LDH) such as D-LDH (gene *dld*) or L-LDH (gene *lldD*) of *E. coli*. We report here the discovery of a novel D- and L-lactate oxidative utilization machinery identified via a comparative genomic reconstruction of *S. oneidensis* MR-1 metabolism combined with physiological, genetic, and biochemical studies.

A hypothetical FeS-containing protein encoded by SO1521 was deemed a candidate for the missing D-LDH based on its presence in the putative operon with an ortholog of lactate permease (SO1522) and its remote homology with the FAD-containing D-LDH from yeast. This prediction was verified by analysis of a SO1521 targeted gene deletion mutant and by genetic complementation of *Escherichia coli* Δdld mutant with a plasmid encoding SO1521. A detailed reconstruction and comparative analysis of lactate utilization subsystem including associated operons and regulons, across hundreds of bacterial genomes integrated in The SEED genomic platform (<http://theseed.uchicago.edu/FIG/>

GTL

subsys.cgi) led to a conjecture that an adjacent three-gene operon SO1518-SO1520 comprised a previously uncharacterized enzymatic complex for the utilization of L-lactate. Two genes of this operon, SO1519 and SO1520 (previously annotated as a hypothetical FeS oxidoreductase and a ferredoxin-like protein, respectively) appear to form a core of this complex conserved in many divergent bacteria (e.g., uncharacterized operons *ykgEF* in *E. coli* and *yvfVW* in *Bacillus subtilis*). This prediction was validated by assay of targeted gene deletions in *S. oneidensis* and by genetic complementation and testing of the L-LDH enzymatic activity in *E. coli* Δdld mutant overexpressing the SO1518-SO1520 operon. Furthermore, the inability of only two of 19 *Shewanella* sp. with completely sequenced genomes to grow with lactate as sole carbon source are consistent with the results of our comparative genome analysis of these species. These findings, in addition to the identification of previously unknown genes involved in lactate utilization in most *Shewanella* species, broadly impact our knowledge of this important aspect of carbon and energy metabolism in many other bacteria. Additional experiments are in progress to elucidate the details of the novel L-LDH complex in *S. oneidensis* and other species. This project is a component of the *Shewanella* Federation and, as such, contributes to the overall goal of applying the genomic tools to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.

66

Multigenome-Based Insights into Respiratory Potential in the Genus *Shewanella*

Margaret Romine^{1*} (margie.romine@pnl.gov), Lee Ann McCue,¹ Gretta Serres,² and Jim Fredrickson¹

¹Pacific Northwest National Laboratory, Richland, Washington and ²Marine Biological Laboratory, Woods Hole, Massachusetts

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.

One of the characteristic features of the Genus *Shewanella* is their remarkable versatility in respiratory metabolism. With the current availability of 17 fully assembled and three partially assembled *Shewanella*

GTL

* Presenting author

genome sequences, it is now possible to predict the breadth of substrates that are utilized and to identify common strategies utilized across this genus to achieve their broad diversity in respiratory metabolism. Each type of respiratory system is expected to include cell envelope-localized proteins which contain metal cofactors (e.g. heme, Fe, NiFe, Mo, FeS, Cu) that mediate electron transfer reactions. In order to develop a global overview of respiratory capabilities in *Shewanella*, we are therefore identifying the suite of redox-active proteins present in these strains, and using them as anchoring points for subsequent neighborhood analysis aimed at identifying all of the cellular components necessary to mediate various respiratory electron transfer functions. By cataloging entire subsystems rather than individual components (e.g. terminal reductases) we are able to assess whether novel components are present relative to systems characterized in other bacterial genera and to identify mutations that may be responsible for subsystem inactivation. Results of these analyses are then used to design experimental studies to test phenotypic predictions or protein function. Results from these experiments in turn are used to refine the computational predictions.

Among the proteins deduced from the genome sequences of 20 *Shewanella*, a total of 138 different types of putative *c*-type cytochromes, encoding between one and 12 CXXCH motifs, and 62 different types of other electron transfer proteins (e.g., ferredoxins, hydrogenases, molybdopterin binding oxidoreductases, formate dehydrogenases, flavoproteins) have been identified. Two extreme outliers in redox protein content were found in this group of species: *S. denitrificans* with only 19 predicted *c*-type cytochromes (overall average=40) and 9 other redox proteins (overall average=22) and *S. sediminis* with 83 predicted *c*-type cytochromes and 35 other redox proteins. Besides being able to grow aerobically, the lone shared respiratory capability among all the sequenced *Shewanella* sp. is the reduction of nitrate to nitrite. In contrast to the other *Shewanella* sp., anaerobic respiratory metabolism in *S. denitrificans* appears to be limited to dissimilatory nitrate reduction to N_2O/N_2 . Occurrence of the dissimilatory reduction to N_2O subsystem in the sequenced *Shewanella* is rare, with most strains instead encoding genes that mediate nitrate ammonification. Common distinctive themes emerging from our analyses include 1) an expanding number of distinct outer membrane deca- and undecaheme *c*-type cytochromes associated with the metal reducing (*mtr*) loci, 2) the occurrence of multiple analogous DMSO reductase-like subsystems comprised of a periplasmic decaheme *c*-type cytochrome, an outer membrane localized molybdopterin oxidoreductase, and an outer membrane protein similar to that found in the *mtr* locus, 3) the occurrence of paralogous respiratory

systems, often in immediate proximity to each other, and 4) the dependence of multiple subsystems on the functionality of a single tetraheme quinol reductase (*CymA*). The large expansion in redox protein content in *S. sediminis* appears to be largely due to the occurrence of paralogous nitrite reductases, DMSO reductase-like systems, and reductive dehalogenases which in part likely reflects its ability to degrade hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX).

67 Identification and Analysis of Components of the Electron Transport Chains that Lead to Reduction of S-Compounds in *S. oneidensis* MR-1

GTL

Eulandria Biddle, Sheetal Shirodkar, and Daad Saffarini* (daads@uwm.edu)

Department of Biological Sciences, University of Wisconsin, Milwaukee, Wisconsin

Project Goals: One of the major goals of the funded research is to identify proteins and enzymes that are involved in the reduction of sulfur-containing compounds in *Shewanella oneidensis* MR-1. In addition, the mechanisms that regulate the reduction of sulfur-compounds will be analyzed. A second goal of the project is to identify proteins that are involved in the regulation of anaerobic respiration in *S. oneidensis*. We have previously identified the cAMP receptor protein as a major regulatory protein of anaerobic respiration. Activation of CRP requires tight regulation of cAMP levels in the cell. One goal of the funded research is to identify phosphodiesterases that may act as oxygen sensors and participate in the regulation of cAMP levels. Additional regulatory proteins that participate in the direct regulation of anaerobic gene expression will also be identified using targeted or transposon mutagenesis. These proteins will be analyzed and their exact functions will be determined.

Shewanella oneidensis MR-1 is able to use S-compounds as electron acceptors for anaerobic respiration. These include tetrathionate, thiosulfate, sulfite, and elemental sulfur. The cAMP receptor protein (CRP) positively regulates the reduction of these compounds in *S. oneidensis*. Two additional regulators are involved in this process. These consist of SO0490 and the two-component system SO4145/SO4147. Mutants that lack SO0490

* Presenting author

exhibit higher levels of sulfur and thiosulfate reduction, but are unable to reduce sulfite. This suggests that this protein plays a dual role in regulating the reduction of S-compounds. SO4144/SO4147 appear to positively regulate the reduction of thiosulfate and sulfur. Interestingly, both SO0490 and SO4145/SO4147 are positively regulated by CRP.

The mechanisms used by *S. oneidensis* MR-1 to reduce S-compounds are not known. It has been suggested that SO4144, an octaheme *c* cytochrome, is a tetrathionate reductase. A mutant that lacks this cytochrome was generated and found to reduce tetrathionate similar to the wild type, indicating that this protein is not the physiological reductase. Mutants deficient in both SO4061 and SO4062 were generated and analyzed. SO4062 is predicted to encode the catalytic subunit of the polysulfide reductase (PsrA), while SO4061 is predicted to encode the Fe-S protein component of the enzyme. Mutants that lack these proteins were deficient in sulfur, thiosulfate, and tetrathionate reduction. These results indicate *S. oneidensis*, unlike other bacteria studied to date, uses the same enzyme to reduce all three compounds.

To identify additional components of the electron transport chain that leads to the reduction of S-compounds, we tested mutants that lack the cytochrome CymA which is involved in electron transfer to Fe(III), nitrate, and fumarate. We also tested mutants that are deficient in cytochrome *c* maturation. Both mutants were able to reduce thiosulfate, sulfur, and tetrathionate similar to the wild type, suggesting that *c* cytochromes are not involved in this process. The polysulfide reductase (PsrABC) responsible for the reduction of sulfur, thiosulfate, and tetrathionate is predicted to be a molybdopterin enzyme that is secreted by the TAT secretion system. This was confirmed by the analysis of mutants deficient in the TAT secretion system and molybdopterin biosynthesis. Both types of mutants were unable to reduce the above-mentioned electron acceptors. Menaquinone deficient mutants were also unable to reduce sulfur, thiosulfate, and tetrathionate. Our results suggest that electrons are transferred directly from menaquinones to the terminal reductase, and additional electron transport components do not appear to be required.

In contrast to thiosulfate reduction, sulfite reduction appears to require *c* cytochromes. Mutants deficient in cytochrome *c* maturation were deficient in sulfite reduction. Similarly, menaquinone biosynthesis mutants were deficient in sulfite reduction. Analysis of additional mutants is ongoing to identify the sulfite reductase and other components that may be involved in the reduction of sulfite.

* Presenting author

“This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.”

68

GTL

Adaptation of *Shewanella oneidensis* MR-1 to its Environment, Insights From Gene Duplication

Margrethe H. Serres^{1*} (mserres@lbl.edu), Daniella Wilmot,¹ and Margaret F. Romine²

¹Marine Biological Laboratory, Woods Hole, Massachusetts and ²Pacific Northwest National Laboratory, Richland, Washington

Project Goals: This project is a component of the *Shewanella* Federation and, as such, contributes to the overall goal of applying the genomic tools to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.

S. oneidensis MR1 is a model organism for studies on anaerobic respiratory metabolism, with particular emphasis on pathways responsible for the biogeochemical transformation of metals and radionuclides. In addition to the diverse respiratory functions encoded by members of the *Shewanella* genus, they also encode a multitude of genes involved in environmental sensing and regulatory responses. These functions are likely important factors that enable *Shewanella* to thrive in aquatic and sedimentary environments all over the world. Genome sequences of 20 *Shewanella* strains are currently available for study of the ecophysiology and speciation of this genus. The ability of *Shewanella* to grow in a range of environments suggests that its metabolic machinery also has undergone an adaptation or specialization to different environmental niches. We are currently identifying the metabolic capabilities of *S. oneidensis* MR1 from using both a bottom up and top down approach that capitalize on the available genome sequence information and on experimental data generated by members of the *Shewanella* Federation. Our goal is to understand how *Shewanella* sp. have adapted to an aquatic lifestyle by identifying common and distinct types of metabolic pathways present in each of the 20 sequenced strains.

We have built a Pathway/Genome Database, ShewCyc for *S. oneidensis* MR-1 using the Pathway Tools software (1). The database is based on the updated annotation of

MR1 provided via ongoing efforts of the *Shewanella* Federation. Gene products have been assigned to metabolic pathways and cellular roles according to the MultiFun classification system embedded in the Pathway Tools software. Over 2800 *S. oneidensis* MR-1 proteins, 60% of the gene products, have been assigned to a metabolic pathway or cellular role. Gene products are given multiple assignments when playing more than one role in the cell. ShewCyc also contains genome information generated by members of the *Shewanella* Federation including sequence similar (paralogous) protein groups, protein complexes, transcriptional units, and regulatory interactions.

Paralogous groups of proteins, arisen through gene duplications followed by divergence, encode related functions. Previous analysis of *S. oneidensis* MR-1 included identifying paralogous groups of proteins encoded in the genome (2). Groups with memberships ranging from 2 to 64 were found. Regulators, transporters, transposases, and chemotaxis-related proteins made up the larger groups, reflecting a diverse ability by *S. oneidensis* MR-1 to sense and respond to its environment. Enzymes were found in groups of 14 or less. By analyzing paralogous enzyme groups, namely those that contain members that function in known metabolic pathways, we are trying to identify proteins representing yet undiscovered metabolic capabilities in *S. oneidensis* MR-1. We selected 756 proteins based on their current assignment to metabolic pathways in ShewCyc. Of these proteins, 459 did not have sequence similar matches in *S. oneidensis* MR-1, while 418 belonged to 141 paralogous groups. All proteins in the 141 paralogous groups were included in the dataset, resulting in a total of 858 proteins being analyzed. The size of the metabolism-associated paralogous groups ranged from 2 to 14, with the largest groups encoding SDR-family oxidoreductases, NAD-dependent epimerases/dehydratases, flavoproteins and enoyl-CoA enzymes. We identified 102 proteins in 56 paralogous groups whose functions are related to known enzymes but whose metabolic role remains to be discovered. A further analysis of these proteins and their occurrence in the other *Shewanella* genomes will be presented.

We would like to acknowledge many members of the *Shewanella* Federation for making their data available for inclusion in the ShewCyc database including Lee Ann McCue and Mary S. Lipton (Pacific Northwest National Laboratory), Mike E. Driscoll and Tim S. Gardner (Boston University), Andrei L. Osterman and Dmitri A. Rodionov (Burnham Institute).

References

1. Karp, P., Paley, S., Romero, P. (2002) The Pathway Tools Software. *Bioinformatics* 18:S225-32.
2. Serres, M.H., Riley, M. (2004) Structural domains, protein modules, and sequence similarities enrich our understanding of the *Shewanella oneidensis* MR-1 proteome. *OMICS* 8:306-21.

69

Cyclic-di-GMP Signaling in *Shewanella oneidensis* MR-1

GTL

Josephine Yu,* Adam Saunders, and Alfred M. Spormann (spormann@stanford.edu)

Stanford University, Stanford, California

Project Goals: Cyclic diguanylate (c-di-GMP) is a new bacterial second messenger that controls diverse cellular processes including biofilm formation, motility, and virulence. *Shewanella oneidensis* MR-1 exhibits the unique capacity to transform multiple organic and inorganic electron acceptors in redox-fluctuating environments. Our goal is to elucidate pathways and molecular mechanisms of c-di-GMP signaling in *Shewanella oneidensis* MR-1 involved in sensing and response to changing redox environments.

Cellular level of c-di-GMP is controlled by the interplay of activities of diguanylate cyclases, which were shown to contain amino acid motifs related to the sequence GGDEF, and specific phosphodiesterases, that often carry an EAL or HD-GYP amino acid sequence motif. Current data indicate that effector proteins, which bind c-di-GMP carry PilZ-like domains. The *S. oneidensis* MR-1 genome encodes more than 60 proteins containing GGDEF and EAL domains, including 19 GGDEF or EAL domain proteins that also carry PAS domains, which are predicted to be involved in redox sensing. Using genomic approaches, we have selected several of these that appear to be physiologically significant and have begun to characterize them biochemically and genetically. Interestingly, expression of one PAS/GGDEF/EAL protein at a low level resulted in a motile phenotype, which switches to a sessile phenotype and a loss of red pigmentation at a higher level of induction. *S. oneidensis* MR-1 possesses five proteins that contain a PilZ domain. Deletion of two of the PilZ-encoding genes severely diminished biofilm formation, and enhanced motility in soft agar assays. Interestingly, deletion of another PilZ-related gene gives the opposite phenotype, resulting in enhanced biofilm formation and

* Presenting author

reduced motility, suggesting an antagonistic role. We are currently determining the mechanism through which PilZ domain proteins transduce the c-di-GMP signal using biochemical and genetic methods.

70

Shewanella baltica: A Model for Examining Specialization Along a Redox Gradient

Jie Deng,* Jennifer M. Auchtung,* Jorge L.M. Rodrigues, and James M. Tiedje (tiedje@msu.edu)

Michigan State University, Center for Microbial Ecology, East Lansing, Michigan

Project Goals: To understand the evolution, ecophysiology, and speciation of different *Shewanella* species.

Shewanella species are thought to play an important role in their ecosystems by coupling the turnover of organic matter with anaerobic respiration of different electron acceptors. Nitrate, an electron acceptor used by many *Shewanella*, is a common contaminant of groundwater as well as several DOE sites (Squillace et al. 2002; Robert and Zachara 1992). One way to remove nitrate from the environment is through the action of denitrifying bacteria. We have begun investigating several strains of *Shewanella baltica* that likely play an important role in the reduction of nitrate. A group of 113 *S. baltica* strains was isolated at different depths from the Baltic Sea, an estuarine ecosystem contaminated with nitrate due to anthropogenic discharges (Ziemke et al. 1998). These *S. baltica* strains represented 77% of the total population of culturable, denitrifying microorganisms. In addition to the isolation of these organisms, a detailed analysis of the water chemistry and the rates of important environmental processes, such as denitrification, were determined throughout the water column (Brettar and Hofle, 1993; Brettar et al. 2001). These analyses revealed that a relatively stable gradient of terminal electron acceptors (oxygen, nitrate, hydrogen sulfide) was present along the depth of the water column.

A subgroup of 37 *S. baltica* strains was selected for further studies. Phylogenetic analysis of the 16S rRNA gene for all 37 strains indicated that they were members of the same species. Thus, in addition to their potential important role in denitrification, these strains also represent an important dataset for investigating short term evolutionary patterns (i.e., divergence among strains from a single geographic location.) A refined phylogenetic

GTL

analysis of this group of strains using a less conserved gene marker, namely the B subunit of the topoisomerase (*gyrB*), identified strain specific differences. Strains were grouped into ten statistically supported clades. This result suggests intraspecies genomic heterogeneity. In order to further investigate this possibility and analyze for genome relatedness, we selected seven genes (SO0578, SO0625, SO1771, SO2183, SO2615, SO2706, SO4702) previously characterized as part of the conserved gene core in *Shewanella* for further studies (Konstantinidis et al. 2006). Specific primers were designed and tested against known sequenced *Shewanella* genomes. Multilocus sequence typing of the selected genes has confirmed seven out of ten clades identified by the *gyrB* sequences. How these genomic patterns translate into their ecological physiology is a central question in biology.

In addition to our phylogenetic analyses, we are also investigating differences in the physiology of these 37 strains. Contrary to the way they were isolated, we were not able to detect denitrification from any of the *S. baltica* strains. Although the strains were not capable of denitrification, we suspect that most of the strains will still play an important role in removing nitrate in the environment by reducing nitrate to ammonia, which can then be consumed by other members of the community. *S. oneidensis* MR-1 reduces nitrate to ammonia (Cruz-Garcia et al. 2007) and the genes required for this process are present in the *S. baltica* isolates that have been sequenced. More physiological tests on these strains are ongoing, specifically we will be testing their growth on different carbon, nitrogen, and phosphorus sources, selection of which will be based on biologic data for two of the *S. baltica* strains. The results of these experiments will highlight physiological differences amongst the strains which, when combined with information about their phylogeny and ecology, should provide insights into the microevolution of this species.

Four *S. baltica* strains were selected for complete genome sequencing based on the following conditions: (1) ecological zone of isolation in the water column (structured as just below the oxic zone with low oxygen and high nitrate (strain OS155), oxic-anoxic transition zone with lower oxygen and high nitrate (strains OS185 and OS223), and the anoxic zone with neither oxygen nor nitrate and low levels of hydrogen sulfide (OS195) and (2) pairwise DNA-DNA reassociation values between 68.2% and 98.2%, representing the same species according to the current bacterial species standards.

Preliminary comparative genomic analysis revealed that although the strains share a large amount of genomic content, the average nucleotide identity between the

* Presenting author

strains ranges from 95.76% (*S. baltica* strains OS155 and OS223) to 97.04% (*S. baltica* strains OS185 and OS195), there are also significant amounts of sequence that vary among the strains. For example, the sequenced genomes of *S. baltica* strains OS155 and OS195 contain 1.1Mb and 1.3Mb of DNA specific to each strain, respectively. These results clearly indicate that these strains are considerably different at the genome level and reveal an unexpectedly large genomic heterogeneity. Experiments are ongoing to determine what effect these differences in genome content have on the cell's physiology and whether they may reflect differential adaptation to microenvironments within the water column. Therefore, through these combined phylogenetic, genomic, and physiologic studies we hope to provide more insight about a major microbial component of a nitrate-contaminated environment while also increasing our understanding of how environment and ecology influence evolution.

References

1. Brettar, I. and M.G. Hofle. 1993. Nitrous oxid producing heterotrophic bacteria from the water column of the central Baltic: abundance and molecular identification. *Mar. Ecol. Prog. Ser.* 94:253-265.
2. Brettar, I., E.R. Moore, and M.G. Hofle. 2001. Phylogeny and Abundance of Novel Denitrifying Bacteria Isolated from the Water Column of the Central Baltic Sea. *Microb. Ecol.* 42:295-305.
3. Cruz-Garcia, C. Murray, A.E., Klappenbach, J.A., Stewart, V., and J.M. Tiedje. 2007. Respiratory Nitrate Ammonification by *Shewanella oneidensis* MR-1. *J. Bacteriol.* 189:656-662.
4. Konstantinidis, K.T., A. Ramette, and J.M. Tiedje. 2007. Toward a more robust assessment of intraspecies diversity using fewer genetic markers. *Appl. Environ. Microbiol.* 72:7286-7293.
5. Riley, R.G. and J.M. Zachara. 1992. Chemical contaminants on DOE lands and selection of contaminant mixtures for subsurface science research. Washington, D.C. U.S. Dept. of Energy, Office of Energy Research. ER-0574T
6. Squillace, P.T., Scott, J.C., Moran, M.J., Nolan, B.T., and D. Koplin. 2002. VOCs, pesticides, nitrate, and their mixtures in groundwater used for drinking water in the United States. *Environ. Sci. Technol.* 36:1923-1930.
7. Ziemke, F., M.G. Hofle, J. Lalucat, and R. Rossello-Mora. 1998. Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. *Int. J. Syst. Bacteriol.* 48 Pt 1:179-86.

71

Linking *Shewanella* Ecophysiology and Molecular Functions

Anna Y. Obraztsova,¹ **Tatiana V. Karpinets**^{2*} (karpinetsv@ornl.gov), Yanbing Wang,¹ Terence B. Kothe,² Denise D. Schmoyer,² Guruprasad H. Kora,² Margrethe H. Serres,³ Margaret F. Romine,⁴ Miriam L. Land,⁵ Edward C. Uberbacher,⁵ and Kenneth H. Nealson¹

¹Department of Earth Sciences, University of Southern California, Los Angeles, California; ²Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; ³Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts; ⁴Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington; and ⁵Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

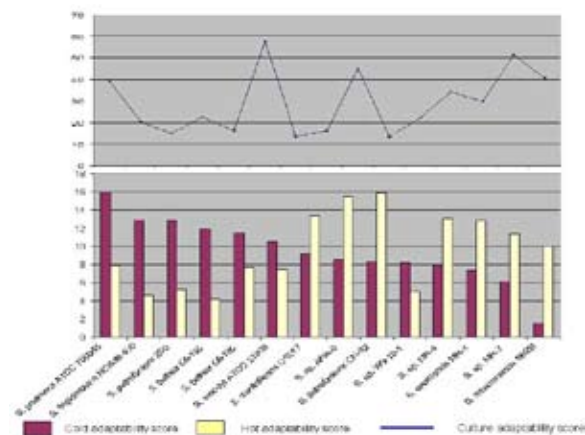
Project Goals: Integrated genome-based studies of *Shewanella* ecophysiology.

Shewanella is an ideal model environmental organism to address a fundamental biological question on linking the adaptive plasticity of the organism to its genotype and to specific cellular processes and molecular mechanisms. Members of the genus inhabit environments that are very different not only in available sources of nutrients and respiration, but also in physical characteristics including temperature, salinity, pH and atmospheric pressure. At present 19 *Shewanella* species from these diverse environments have been sequenced providing an opportunity for comprehensive comparative analysis of their genomes. However, linking this genomic information to complex physiological traits, like bacterial adaptation to culture conditions or to cold and hot temperatures, is challenging, because qualitative characterizations of such complex traits are usually absent.

Objective of the study was to characterize and to relate adaptability of the sequenced *Shewanella* species to different culture conditions to their genomic characteristics. The approach was based on quantifying growth phenotypes and relating them to molecular functions encoded by the genomes. The adaptive plasticity of 15 sequenced *Shewanella* species were characterized experimentally by their growth profiling in temperatures ranging from 0 to 55 degrees C. Bacteria were cultivated aerobically using minimal SF medium with lactate as carbon source. LB medium and marine broth were used only for *S. denitrificans* and *S. woodyi* respectively. The optical density

* Presenting author

(OD₆₆₀) measurements were used to quantify the adaptive potentials of each species. Specifically, the adaptability to hot temperature was estimated by the Hot Adaptability Score. The score was calculated as the percentage of the accumulated ODs at 30–37 °C if compared with the ODs accumulated at 24 °C, which was optimal for most species. The Cold Adaptability Score calculated as the percentage of the accumulated ODs at 4–12 °C. The Culture Adaptability Score was calculated as the accumulated ODs at optimal growth temperature. For all calculated scores, the greater the score, the greater is the adaptability of the species. This quantification of the analyzed phenotypic traits allowed us to compare the adaptive potentials of the species to culture conditions, cold and hot temperatures, and then to relate these potentials to the enrichment of each genome with different molecular functions, which were characterized in terms of protein families as determined by Pfam. To characterize molecular functions involved in transport and proteolysis in more detail the genomes were also annotated using TransportDB and the peptidase database MEROPS. The number of domains representing each molecular function in each genome was calculated using SQL queries.



Results of the growth profiling (Figure) indicate significant difference in the adaptability of *Shewanella* to temperature and culture not only across species, but also across strains of the same species. The species with a better adaptation to cold temperatures tend to have poorer adaptability to hot temperatures; and the adaptation to culture conditions doesn't correlate to thermal adaptation of the species. These experimental observations indicate that different molecular functions underlie the studied physiological traits. We found that the culture adaptability had the greatest number of molecular functions whose enrichments significantly correlated with the score across species. Most known identified domains represented enzymes involved in proteolysis, including peptidases of different MEROPS families. Another group of the correlated domains revealed the importance of some specific

repeats and transposases for successful adaptation of the *Shewanella* genome to culture conditions. Several identified domains represented molecular functions involved in the bacterial type II and III secretion systems and key elements in two-component signal transduction systems. This project is a component of the *Shewanella* Federation and contributes to revealing molecular functions underlying the diverse ecophysiology of this important species.

“The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.”

72 Motility and Chemotaxis in *Shewanella oneidensis* MR-1

GTL

J. Li,¹ W. Song,¹ and M.J. Ward^{1*} (mjward@jhu.edu)

¹Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, Maryland

Project Goals: Characterize the ecophysiology of *Shewanella oneidensis* by understanding the motility and chemotaxis systems.

Shewanella are a group of microbes that show extreme respiratory versatility. These microbes are also usually found in redox-stratified environments. Taken together, these characteristic features suggest that *Shewanella* ecophysiology is tightly linked to their ability to sense chemical gradients in their environment and respond appropriately. Our role, as members of the *Shewanella* Federation, is to understand how and why these responses occur by analyzing the motility and chemotaxis systems of the model organism, *S. oneidensis* MR-1.

By rotating a single polar flagellum, either clockwise or counter-clockwise, *Shewanella* swim as a series of forward movements and reversals. *S. oneidensis* MR-1 cells have been tracked moving at over 100 μm/sec and with an unstimulated reversal frequency of 0.6 – 0.9 reversals/sec. The flagellar machinery strongly resembles that found in *E. coli*, and ‘late’ flagellar gene expression is similarly regulated by the alternative sigma factor, sigma 28 (σ^{28}). Using an iterative position specific score matrix-based approach, we have identified several genes that are predicted to have σ^{28} promoters and that are part of a σ^{28} -regulated transcriptional network based on

* Presenting author

transcriptional profiling data from the Shewy Correlation Browser (Gardner Lab, Boston University). The proteins encoded by these genes are not found in *E. coli* and thus may be *Shewanella*-specific components of the motility apparatus.

The chemotactic machinery of *S. oneidensis* MR-1 appears substantially more complex than that of *E. coli*. However, deletion of a single gene, *cheA-3*, results in a non-chemotactic phenotype. Studies involving this non-chemotactic mutant have shown that *S. oneidensis* responds to both electron donors/carbon sources and electron acceptors. Our initial hypothesis, that receptors with redox-sensing PAS domains would dominate behavioral responses by monitoring electron flux through the respiratory electron transport chain, appears to be incorrect, although a Δ SO3404 mutant does show modified aerotactic behavior. Interestingly, another receptor, that we predicted would be required for responses to *N*-acetyl glucosamine, appears to actually be involved in preventing movement towards high concentrations of this electron donor/carbon source. The detailed results of our studies will be presented at the meeting.

We would like to note that conversations with many members of the *Shewanella* Federation have contributed to, or provided direction for, much of our research. We are particularly indebted to Margie Romine (PNNL), Tim Gardner (Boston University), and Igor Zhulin (ORNL).

73

GTL

Integrated Genome-Based Studies of *Shewanella*

Haichun Gao,^{1,2} Jingrong Chen,^{1*} Yili Liang,¹ Zhili He,^{1*} Margaret Romine,³ Kenneth Nealon,⁴ James M. Tiedje,² James K. Fredrickson,³ and Jizhong Zhou¹ (jzhou@ou.edu)

¹University of Oklahoma, Norman, Oklahoma;

²Michigan State University, East Lansing, Michigan;

³Pacific Northwest National Laboratory, Richland,

Washington; and ⁴University of Southern California, Los Angeles, California

Project Goals: The overall goal of this project is to apply the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems we propose to use genome-based approaches to investigate *Shewanella* as a system of integrated networks; first

describing key cellular subsystems—those involved in signal transduction, regulation, and metabolism—then building towards understanding the function of whole cells and, eventually, cells within populations.

Shewanella is comprised of more than 40 recognized species inhabiting a wide range of niches planet-wide including the terrestrial subsurface, wetlands, redox interfaces in marine and freshwaters, and cold waters and sediments of the deep sea. Most of these species can utilize multiple electron acceptors including O₂, NO₃⁻, S₀, Mn^{+3,4}, Fe⁺³, Cr⁶⁺, V⁵⁺, DMSO, TMAO, and fumarate. The ability of *Shewanellae* to respire a wide variety of electron acceptors, use the products of other microorganisms and biopolymers such as chitin, and to sense and taxis towards electron acceptors make *Shewanellae* well adapted to living in organic-rich, redox gradient environments. As a part of the *Shewanella* Federation efforts, we have integrated genomic technologies to study various aspects of energy metabolism of two *Shewanella* strains from a systems-level perspective.

Nitrate reduction in *Shewanella oneidensis* MR-1.

S. oneidensis MR-1 serves as a model for studying anaerobic respiration and electron transport-linked metal reduction. In the genome of *S. oneidensis*, a *napDAGHB* gene cluster encoding periplasmic nitrate reductase (NapA of the NAP system) and accessory proteins and an *nrpA* gene encoding periplasmic nitrite reductase (NrpA of the NRF system) have been identified. However, these two systems appear to be atypical because the genome lacks both *napC* and *nrpH*, which are essential for reduction of nitrate to nitrite and nitrite to ammonium in most bacteria containing these two systems, respectively. In this study, we demonstrated that reduction of nitrate to ammonium in *S. oneidensis* is carried out by these atypical systems in a two-step manner. Unexpectedly, the *napB* mutant exhibited a higher maximum cell density than the wild-type while the *napA* mutant was defective completely in growth on nitrate. Although reduction of nitrate to ammonium in the *napB* mutant is also conducted by NAP and NRF systems, nitrite, the intermediate of the reduction, was not detected through the entire reduction. Further investigation suggests that NapB may be the preferred electron acceptor from a membrane-bound protein which delivers electrons from metaquinol pool to a number of terminal reductases. In an attempt of searching for this membrane-bound protein, both microarrays and mutational analysis have been taken. Results suggest that CymA is likely to be functional replacement of both NapC and NrpH in the nitrate reduction and a novel conceptual model for nitrate reduction is proposed.

* Presenting author

Genomic Array Footprinting in *Shewanella oneidensis* MR-1. Genomic Array Footprinting (GAF) is a high-throughput method to identify conditionally essential genes in microbes by using a combination of random transposon mutagenesis and microarray technology. A GAF was developed for *S. oneidensis* by constructing plasmid pJZ214 from widely used pBSL180. An *S. oneidensis* MR-1 transposon insertion library was generated with pJZ214, containing clones representing $\sim 10^5$ independent insertions. This insertion library was subjected to a competitive growth selection in minimal medium with either lactate or glycyl-glutamate as only carbon source for 67 generations in triplicate. Cells were sampled daily. Genomic DNAs from original library and samples were extracted and digested by *DpnI* to generate DNA fragments of an average size of 400 bp. These fragments were then used as templates for *in vitro* transcription by T7 RNA polymerase and the resulting RNA was labeled by reverse transcription. The labeled cDNAs from samples and original library were cohybridized onto *S. oneidensis* whole-genome microarrays. Meanwhile, conventional microarray was conducted with mRNA from exponential phase cells grown on lactate or glycyl-glutamate. Microarray data from conventional microarray showed that nearly half of genes encoding predicted peptidases (ie. *dcp-1*, *so1075*, *so0614*) and a few of genes encoding peptide transporters were up-regulated by the presence of glycyl-glutamate. GAF results from samples with either carbon source revealed a similar number of (~ 450) genes with reduced signals compared to the original library. These genes belong to a variety of functional categories, suggesting that growth on single carbon source is still a complicated biological process. Although the GAF results were generally consistent with data from conventional microarray analysis, discrepancies were also found. To find the selection pattern of the insertions with respect to the genes' functional classification, systematic analysis is undertaken.

Metal reduction in *S. putrefaciens* W3-18-1. *S. putrefaciens* exhibits an extraordinary ability to reduce irons in various forms. However, little is known about its molecular basis. Predicted counterparts in W3-18-1 to *omcA* and *omcB* of *S. oneidensis* MR-1 are *sputw3181-2445* and *sputw3181-2446*, respectively. Unlike MR-1, W3-18-1 lacks homologs to MtrDEF, which are believed to be composed of a secondary metal reductase. To examine whether *Sputw3181-2445* and *Sputw3181-2446* of W3-18-1 are functionally equivalent to *OmcA* and *OmcB* in MR-1, in-frame *sputw3181-2445* and *sputw3181-2446* individual deletion and double deletion strains were constructed and their metal reduction characteristics were examined. When grown under anaerobic conditions in LB medium containing 20mM sodium

lactate as the electron donor and one of following as the electron acceptor: ferric citrate, MnO_2 , V_2O_5 , MoO_3 , Cobalt(III)-EDTA, α -FeOOH, β -FeOOH, Fe_2O_3 or $Fe(OH)_3$, reduction characteristics of these mutants resemble their counterparts of MR-1. These results suggest that *OmcA* and *OmcB* in W3-18-1 functions similarly to those in MR-1. Interestingly, the residual metal reduction ability remains even in the *sputw3181-2445* and *sputw3181-2446* double mutant of W3-18-1. A similar observation in MR-1 was presumably due to the presence of MtrDEF. On the basis of that homologs to MtrDEF are missing in W3-18-1 our findings suggest a possible existence of additional unknown metal reductase(s).

74

GTL

Mechanisms of Sulfur Reduction by *Shewanella*

Karlo Lopez, Vinita Lukose, Sneha Vakimudi, and Edward (E.J.) Crane* (EJ.Crane@pomona.edu)

Department of Chemistry, Pomona College, Claremont, California

Project Goals: The focus of our group, within the *Shewanella* Federation's overarching goal of understanding *Shewanella*'s ecophysiology, is to determine the pathways of sulfur reduction, and to determine the detailed mechanism of sulfur/polysulfide reduction at the molecular level. Our efforts are divided into 3 main areas of emphasis, which are 1) studies of the native *S. oneidensis* MR-1 sulfur reduction pathway and the native PsrABC complex, 2) construction of recombinant PsrABC complex, produced via heterologous overexpression, for the purpose of detailed mechanistic studies, including site-directed mutagenesis and behavior in immobilized films, and 3) detailed characterization of a recently discovered soluble polysulfide reductase present in *S. lobica* PV-4 and *S. frigidimarina*.

The ability to engage in dissimilatory sulfur reduction (DSR) is a general characteristic of the facultatively anaerobic genus *Shewanella*. While previously thought to be limited to strict anaerobes, the ability to grow while using sulfur as an electron acceptor fits in with the tendency of *Shewanella* to use a wide range of electron acceptors. All of the sequenced genomes of *Shewanella* species contain homologues to the subunits of the polysulfide reductase complex (PsrABC; one *Shewanella* member – *S. denitrificans* – does not contain PsrABC homologues and cannot perform DSR) that has been previously purified from the strictly anaerobic sulfur

* Presenting author

reducer *Wolinella succinogenes*. The basic characteristics of the *W. succinogenes* complex have been deduced in previous studies, however, because of the multiple coenzymes and membrane-bound nature of the complex a detailed understanding of the mechanism of the complex remains to be determined, as do the identities of the direct electron donors, protein or otherwise, to the complex.

The focus of our group, within the *Shewanella* Federation's overarching goal of understanding *Shewanella*'s ecophysiology, is to determine the pathways of sulfur reduction, and to determine the detailed mechanism of sulfur/polysulfide reduction at the molecular level. Our efforts are divided into 3 main areas of emphasis, which are 1) studies of the native *S. oneidensis* MR-1 sulfur reduction pathway and the native PsrABC complex, 2) construction of recombinant PsrABC complex, produced via heterologous overexpression, for the purpose of detailed mechanistic studies, including site-directed mutagenesis and behavior in immobilized films, and 3) detailed characterization of a recently discovered soluble polysulfide reductase present in *S. lobica* PV-4 and *S. frigidimarina*.

The pathway of electrons from donors to their ultimate destination on sulfur is not currently known. Through extensive analysis of mutants, we have shown that in *S. oneidensis* MR-1 DSR is not likely to use any of the multitude of cytochromes or cytochrome type-proteins (including the "Mtr" proteins) present in the organism. A mutant deficient in menaquinone synthesis, however, does indicate that quinone synthesis is necessary for DSR to occur. We have been able to use membrane extracts to reproduce the sulfur reduction pathway in a cell-free system using both lactate and hydrogen as electron donors. We are currently surveying electron acceptors that will allow us to purify the PsrABC complex (the complex is usually assayed via the reverse reaction, the reduction of compounds by S^2), since further purification of the complex is likely to remove proteins such as hydrogenase that act in the DSR-electron transfer pathway.

In the construction of the recombinant complex, we have overexpressed and purified the subunit that serves as the membrane anchor (psrC), which is purified with the expected bound quinone, suggesting that the soluble overexpressed protein is in an active form, despite its *in vivo* nature as an integral membrane protein. Some of this surprising solubility may be derived from the thioredoxin tag present at the N-terminus of the protein during expression and purification, although the protein remains soluble after the thioredoxin (and poly-histidine) tag is cleaved from the protein. The PsrB subunit has also been overexpressed and purified, and shows UV-visible

spectra consistent with the presence of the Fe-S centers observed on the *W. succinogenes* psrB protein. When the PsrC subunit was incubated with a polyanionic film, QCM results indicate that the protein was absorbed into the film, essentially sinking down "into" the film in a manner analogous to the interaction of an integral membrane protein with a membrane. When PsrB was incubated with the PsrC-containing film, the results were consistent with PsrB binding on the surface of the film or the PsrC protein. Both PsrB and C were resistant to being washed off of the film by buffer, and appear to be essentially immobilized within or on the film. We are currently in the process of cloning and overexpressing the PsrA subunit, which contains the molybdopterin site that is likely to be the site of polysulfide reduction. PsrA has been proposed to behave similarly to PsrB in terms of its interaction with PsrC, and the behavior of the PsrB and C subunits (which we had expected to be the most troublesome of the system) inspire confidence that reconstitution of active PsrABC immobilized in a polyanionic film will be feasible, providing a convenient and reproducible system for the in-depth analysis of the redox mechanism of the complex and its molybdopterin center.

During our survey of *Shewanella* genomes we noticed that *S. lobica* PV-4 and *S. frigidimarina* both contained homologues to a family of flavoproteins that our lab had previously studied in the hyperthermophile *Pyrococcus*, although the *Shewanella* protein had an additional "tail" homologous to the polysulfide carrier protein (Sud) from *W. succinogenes*. Results from another laboratory demonstrated that proteins in this family can be involved in DSR (specifically in the reduction of sulfur), and we found that the overexpressed and purified FAD-dependent protein from *S. lobica* PV-4 acts as an NADH-dependent and CoenzymeA-activated polysulfide reductase. This is an entirely new reaction, since the previously observed reaction with the *Pyrococcus* enzyme was elemental sulfur reduction in a CoA-dependent manner, with the NADH substrate showing an extremely high K_m (5 mM), while the reaction we've observed is the reduction of polysulfide in a CoA-activated manner, with a low μ molar K_m for NADH. We are currently in the process of characterizing the mechanism of this enzyme using site-directed mutagenesis and steady-state and presteady-state kinetic techniques.

* Presenting author

75

GTL

Coupled Informatic-Experimental Analyses of Carbon Metabolism Subsystems in *Shewanella*

G. Pinchuk,^{1*} O. Geydebekht,¹ D. Kennedy,¹ S. Reed,¹ M. Romine,¹ C. Yang,² D. Rodionov,² Andrei Osterman,² J. Reed,³ J. Scott,⁴ **A. Beliaev**¹ (alex.beliaev.pnl.gov), and **J. Fredrickson**¹ (jim.fredrickson@pnl.gov)

¹Pacific Northwest National Laboratory, Richland, Washington; ²Burnham Institute for Medical Research, La Jolla, California; ³University of Wisconsin, Madison, Wisconsin; and ⁴Dartmouth College-Earth Sciences, Hanover, New Hampshire

Project Goals: *Shewanella oneidensis* MR-1 is a motile, facultative γ -Proteobacterium with remarkable respiratory versatility; it can utilize a range of organic and inorganic compounds as terminal electron acceptors for anaerobic metabolism. More broadly, *Shewanellae* are recognized free-living microorganisms and members of microbial communities involved in the decomposition of organic matter and the cycling of elements in aquatic and sedimentary systems. To function and compete in environments that are subject to spatial and temporal environmental change, *Shewanella* must be able to sense and respond to such changes and therefore require relatively robust sensing and regulation systems. The overall goal of this project is to apply the tools of genomics to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems we propose to use genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems and then building towards understanding the function of whole cells and, eventually, cells within populations. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments.

The genus *Shewanella* is unusually well-adapted to chemically (redox) stratified environments as reflected in the ability to utilize a broad range of electron acceptors via a highly diversified electron transport system. Occupying such niches requires the ability to adapt rapidly to changes in electron donor/acceptor type and availability; hence the ability to compete and thrive in such environments must ultimately be reflected in the

organization and flexibility of the central carbon metabolism pathways. Although *Shewanella* species are typically considered to have a relatively restricted substrate range for carbon and energy sources, genome-based analyses revealed multiple pathways for C₂₋₆ compounds, amino acids, and fatty acids, reflecting its ecological role as a consumer of organic matter breakdown products in relatively carbon-rich environments that support diverse anaerobic microbial communities. Using controlled cultivation, biochemical, genetic and genomic approaches in conjunction with pathway modeling, we showed that (i) lactate utilization employ previously unknown novel enzymes; (ii) metabolic pathways expressed under different redox conditions utilize pyruvate as a key metabolite and (iii) the pathways involved in ATP production under aerobic and anaerobic conditions fundamentally differ reflecting the amount of energy this organism can generate by oxidative phosphorylation.

Novel enzymes for L- and D-lactate utilization. The first and essential step of lactate utilization is its oxidation to pyruvate, a reaction catalyzed by diverse group of lactate dehydrogenases. The genome of *S. oneidensis* MR-1, the subject of detailed annotation, revealed only one candidate gene, SO0968, predicted to be an NAD⁺-dependent lactate dehydrogenase (LDH). Using genetic and biochemical approaches, we found that SO0968 product does not catalyze lactate oxidation but rather encodes a NADH:pyruvate reductase that is involved in pyruvate fermentation and can be used to generate energy in the absence of an electron acceptor. Further research using subsystems-based comparative genomic analysis uncovered novel D- and L-lactate oxidative utilization machinery.

Metabolic pathways expressed under different redox conditions utilize pyruvate as a key metabolite. One of the unresolved problems of *Shewanella* biology is the inability of this organism to couple anaerobic respiration to the oxidation of compounds other than lactate and pyruvate. Growth tests with different organic compounds confirmed that MR-1 can use acetate, succinate, α -oxoglutarate, and propionate as a sole source of carbon and energy under aerobic, but not anaerobic, conditions with fumarate and Fe(III) as electron acceptors. To investigate the underlying mechanisms involved in anaerobic oxidation of carbon substrates, we carried out a comparative analysis of *S. oneidensis* MR-1 metabolism under aerobic, O₂-limited, and anaerobic conditions using integrated genetic, controlled cultivation and functional genomics approaches. Collectively, the results indicate that a shift from aerobic to O₂-limitation leads to fundamental changes in gene expression and enzyme activities that re-wire cells for anaerobic growth.

* Presenting author

Under O_2 -limited conditions, pyruvate oxidation was catalyzed exclusively by pyruvate-formate lyase (locus tag: SO_2912), more than 80% of metabolized lactate was converted to acetate through phosphate acetyltransferase and acetate kinase, and biomass yield decreased about 2.5-fold compared to aerobic growth. These changes were accompanied by increased expression of anaerobic respiratory genes including those involved in fumarate, DMSO, and metal reduction. While the full TCA cycle was indispensable for aerobic growth on all compounds tested, the inactivation of α -oxoketoglutarate activity did not affect the anaerobic growth of MR-1 with lactate and fumarate/Fe(III). Moreover, a deletion of the E1 subunit of pyruvate dehydrogenase abolished the ability of MR-1 to grow anaerobically with any of the substrates tested including acetate. It did not, however, impair the ability to grow anaerobically with lactate or pyruvate. In contrast, inactivation of pyruvate-formate lyase did not affect aerobic MR-1 metabolism but impaired anaerobic growth as well as pyruvate fermentation. Overall, our results strongly suggest that oxidation of all tested substrates in *S. oneidensis* MR-1, including acetate, proceeds through pyruvate making this compound a very important oxidative intermediate. The presence of an incomplete TCA cycle and a tendency to oxidize electron donors through pyruvate is likely a key reason why *Shewanella* is unable to use α -ketoglutarate and acetate under anaerobic conditions.

The role of substrate-level phosphorylation under anaerobic and O_2 -limited growth of *S. oneidensis* MR-1.

One of the fundamental characteristics of *S. oneidensis* metabolism is its inability to use acetate as an electron donor under anaerobic conditions. We extended these previous observations by demonstrating that acetate also cannot be used as an energy source under anaerobic nor O_2 -limited conditions. Anaerobic or O_2 -limited growth with lactate as the electron donor is accompanied by acetate excretion. We hypothesized that acetate excretion is coupled to ATP production which is catalyzed by acetate kinase (SO_2915). In support of this hypothesis, an MR-1 acetate kinase mutant did not grow anaerobically with either Fe(III)-NTA or fumarate when lactate served as the carbon and energy source. Chemostat experiments also showed that the amount of acetate produced was in inverse proportion to the O_2 flux suggesting that under these conditions, *S. oneidensis* MR-1 growth depends solely on substrate level phosphorylation. We have generated several lines of evidence, including flux balance analysis of *S. oneidensis* metabolism, showing that under conditions of O_2 limitation or fumarate reduction most of ATP is produced from lactate or pyruvate via substrate level phosphorylation (from acetyl phosphate). Our results suggest that, for *S. oneidensis*, the rate of electron

transfer to a terminal electron acceptor determines the growth rate and fraction of energy spent on maintenance needs, whereas efficiency of electron transport coupling to phosphorylation partially determines biomass growth yield. Such flexibility of central carbon metabolism would allow *Shewanella* to survive during periods of nutrient-limitation and proliferate rapidly when both electron acceptor(s) and donor(s) are available.

76

GTL

Metabolic Reconstruction of *Shewanella oneidensis*: A Community Resource

Jennifer L. Reed^{1*} (reed@engr.wisc.edu), Iman Famili,² Sharon J. Wiback,² Grigoriy Pinchuk,³ Margaret R. Romine,³ Johannes C. Scholten,³ Joel Klappenbach,⁴ and James K. Fredrickson³

¹University of California, San Diego, Department of Bioengineering, La Jolla, California; ²Genomatica, Inc, San Diego, California; ³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington; and ⁴Center for Microbial Ecology, Michigan State University, East Lansing, Michigan

Project Goals: This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.

Genome-scale network reconstructions account for components and component interactions in biological networks, and are a way in which to collate and analyze data from a variety of sources. Here we report a metabolic reconstruction for *Shewanella oneidensis* MR-1 based on the current genome annotation and primary literature. The reconstruction includes 761 reactions, 789 genes, and 625 metabolites.

The reconstruction was used to build a flux balance model that was used in a variety of computational analyses, including: assessment of growth phenotypes, evaluation of metabolite usage (as substrates or by-products), and prediction of knock-out phenotypes to look at metabolic robustness. The model correctly predicted growth on a variety of carbon and nitrogen sources. In addition, quantitative evaluation of alternative electron acceptors led to the identification of 7 classes of electron acceptors, with differing biomass yields (g D.W. produced per

* Presenting author

mmol electron acceptor consumed). Gene deletion simulations across 10 different environmental conditions with various carbon sources and electron acceptors found that a large fraction of genes were never essential (542 out of 779), while a smaller fraction were always essential (198 out of 779) for growth on these 10 conditions.

Together this work provides a resource that can be used by *Shewanella* researchers and illustrates how reconstructions can serve as a means to evaluate experimental data and generate testable hypotheses to better understand its ecophysiology. This project is a component of the *Shewanella* Federation and as such contributes to the overall goal of applying the tools of genomics to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus.

77

GTL

The *Shewanella* Knowledgebase

Ed Uberbacher^{1*} (uberbacherec@ornl.gov), Nagiza Samatova,^{1,2} Tatiana Karpinetz,³ Erich Baker,⁴ Guru Kora,³ Andrey Gorin,¹ Mike Leuze,¹ Mustafa Syed,¹ and Denise Schmoyer¹

¹Oak Ridge National Laboratory, Oak Ridge Tennessee; ²North Carolina State University, Raleigh North Carolina; ³University of Tennessee, Knoxville, Tennessee; and ⁴Baylor College of Medicine, Waco, Texas

.....
<http://shewanella-knowledgebase.org>

The *Shewanella* Knowledgebase is designed to provide a framework for Federation investigators to share, combine and analyze data. The first version of this database was released to the Federation earlier this year. In the last six months or so the effort has focused on a number of issues designed to make the resource more valuable to Federation investigators. These include (i) improving the coverage and support for the many data types generated in the SF, (ii) improving the data linkage to key investigators and developing procedures to capture their data streams, (iii) developing database support for multiple *Shewanella* species and strains, and (iv) developing improved linkage to other data resources in the SF, community and to reference materials, (v) collecting information on the species available in literature, internet resources and databases, (vi) integrating the collected information in the analysis of the experimental data produced by SF.

The project has been successful at defining procedures to keep up with data production at most SF labs. Support

for new types of data and projects, including physiology experiments and biolog data have been implemented. These include new capabilities for graphical data associated with experiments, growth curves, and images.

Support for comparative analysis of multiple species has been implemented, including the construction of Shew-Cyc pathway databases. Tools for species comparison at pathway and genome levels are available and improving. Regulatory data has been integrated from numerous sources as a component of the Knowledgebase. Computational predictions of the regulatory elements in the bacteria were collected from the published literature and from the different Internet resources including Rfam, RibEx, TractorDB, RegTransBase, BioCyc, PromScan, and others. This information was analyzed to identify a set of the basic regulatory classes for their structural presentation in the relational database. They include translated coding sequences, DNA regulator binding sites, sigma factor binding sites, transcription units, promoters, regulons, stimulons, and RNA regulators. The last category encompasses a diverse class of regulators including non-coding and small RNAs, different types of terminators and riboswitches. This component of the Knowledgebase is described in a companion abstract.

Greatly improved linkage to numerous SF and community data sources has been established. A publication mining system that includes a master table of publications with links to the reference sources, authors, and Knowledgebase projects and text search system has also been implemented. Procedures to maintain and update this library are in place. The resource now has *Shewanella* member information, their contributions in terms of data, publications and literature, and a universal resource locator for personal contact.

Web Portal. The *Shewanella* Knowledgebase web portal is a data and knowledge integration environment that allows investigators to query across the *Shewanella* Federation experimental datasets, link to *Shewanella* and other community resources, and visualize the data in a cell systems context. The web portal has many intuitive ways of exploring federation experimental data. The data can be downloaded to the user's computers in the original format. Various data navigation features are also presented to explore the data on the server. *Shewanella* database backend is coupled with powerful system-wide search feature that includes all federation data, publications, literature, etc. The *Shewanella* front-end is built using a combination of Web 2.0 presentation layer technologies. The web portal is built with HTML 4.0, CSS & Script.aculo.us javascript library. The content is mostly

* Presenting author

dynamically generated using Java Server Pages Standard Tag Library (JSTL).

Data Analysis. The user interface has many intuitive guides and wizards to explore *Shewanella* experimental results. It features many experimental data comparative analysis modules that perform one-on-one analysis with diverse sets of biological data, with corresponding visualization capabilities at various data aggregation levels and in different biological contexts. It also provides a unified set of integration analysis tools that currently support ShewCyc pathways and pathway group categories. Future releases will include KEGG pathways, TIGR roles, and GO ontologies for exploring data.

Data Visualization. Various data visualization schemes are provided to display the results of *Shewanella* Federation experiments. One of the viewers compares relative expression data at the gene level, while other viewers compares the average or percentages of the under/over expressed genes in a pathway or pathway group. These viewers are also cross-referenced to Pathway Tools software which contains reference pathways for multiple *Shewanella* strains.

Computing Infrastructure. The computing infrastructure for the database has been further developed. The *Shewanella* web portal has a multi-tier architecture with each tier hosted on a separate server and designed to work independently. It is build with open source MySQL database and is hosted on a power/data redundant server at ORNL. The *Shewanella* backend database provides numerous security features like row level locking, user level access privileges etc.

“The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.”

78

Comprehensive Integration of Regulatory Data in the *Shewanella* Knowledgebase

Mustafa H. Syed,¹ Tatiana V. Karpinets,² Denise D. Schmoyer,² Michael R. Leuze,² Nagiza Samatova,^{2,3} and Edward C. Uberbacher^{1*} (uberbacherec@ornl.gov)

¹Biosciences Division and ²Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee and ³North Carolina State University, Raleigh, North Carolina

Project Goals: Data integration and sharing for the *Shewanella* Federation.

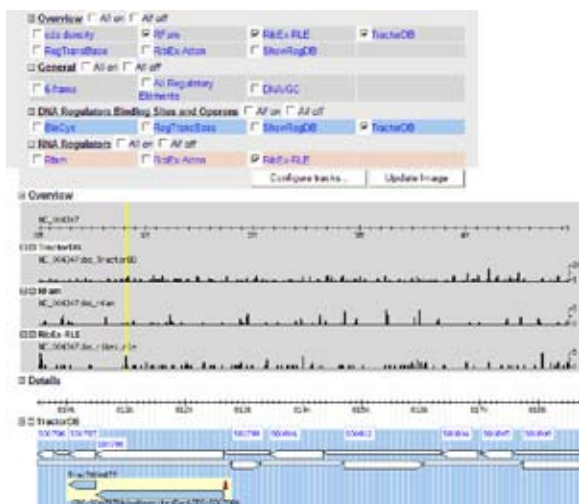
It has become increasingly evident over the recent years that cell regulation is a complex multilayer process encompassing all steps of the information processing in the cell and including a diverse set of regulatory elements. Untranslated regions in *Shewanella* may be considered as a complex regulatory system that can both produce and integrate the regulatory information in the cell, thereby adjusting cellular processes to the environment. Experimental identification of the regulatory elements, however, is very laborious, and as a result numerous computational methods have been developed to predict elements. even in the model organisms it involves a lot of computational predictions. The results of many such predictions for *Shewanella* are found in many different regulation-related databases. Thus, highly relevant regulatory information related to *Shewanella* is scattered across publications and the different Internet resources. At present there is no accepted infrastructure that allows one to integrate, visualize, and analyze the regulatory information based on different computational predictions for a particular organism. This infrastructure, however, is very important to gain insight into *Shewanella* regulation at the level of each individual gene, to check the predictions and to integrate the diverse regulatory information in the analysis of experimental data, especially those obtained by large-scale technologies, like microarray or proteomics.

Objective of this work was to develop an infrastructure for collection, storage, and visualization of the regulatory information relevant to *Shewanella oneidensis* MR-1 and to integrate this information with numerous experimental datasets submitted to the *Shewanella* Knowledgebase (<http://shewanella-knowledgebase.org>). The computational predictions of the regulatory elements were collected from the published literature and from

* Presenting author

the different Internet resources including Rfam, RibEx, TractorDB, RegTransBase, BioCyc, PromScan, and others. This information was analyzed to identify a set of the basic regulatory classes and define their representation in the Knowledgebase. Elements include translated coding sequences, DNA regulator binding sites, sigma factor binding sites, transcription units, promoters, regulons, stimulons, and RNA regulators. The last category encompasses a diverse class of regulators including non-coding and small RNAs, different types of terminators and riboswitches.

Visualization of the collected *Shewanella* regulatory information was implemented using Gbrowse from the Generic Model Organism Database Toolkit (<http://iubio.bio.indiana.edu/gmod/gbrowse/>), which was configured for the *Shewanella oneidensis* MR-1 genome and adjusted to a specificity of the collected information (see Figure). Options currently available in the regulatory element browser include (a) an overview of all regulatory elements in the genome with scrolling to any selected region, (b) presentation of each type of element on different tracks or all elements together on one track, (c) different types of zooming, d) simultaneous bird's eye and detailed views of the genome, (e) brief information on each element by popup balloons and (e) detailed information on each regulatory element including decorated (colored conserved regions) FASTA sequences. Users can download regulatory sequences and feature tables in various formats.



“The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.”

* Presenting author

79 Adaptive Evolution of *Geobacter* to Optimize Electricity Production and Bioremediation and to Elucidate Complex Physiological and Ecological Phenomena

Zara Summers, Kelly P. Nevin* (knevin@microbio.umass.edu), Shelley A. Haveman, Hana Yi, Trevor L. Woodard, and **Derek R. Lovley** (dlovley@microbio.umass.edu)

University of Massachusetts, Amherst, Massachusetts
 Project Goals (Abstracts 79, 84-86): The goal of this project is to provide computational tools to predictively model the behavior of two microbial communities of direct relevance to Department of Energy interests: 1) the microbial community responsible for in situ bioremediation of uranium in contaminated subsurface environments; and 2) the microbial community capable of harvesting electricity from waste organic matter and renewable biomass. Based on our current rate of progress, it is expected that within five years it will be possible to predict the in situ growth and activity of the relevant microorganisms in the environments of interest solely from relevant geochemical data and to describe in detail the in situ metabolic state of the microorganisms from environmental gene expression data. Furthermore, these computational tools will be able to predict the response of the microbial community to environmental manipulations or manipulation of the genome of the relevant organisms, allowing rational optimization of in situ uranium bioremediation or electricity harvesting via environmental or genetic engineering.

Geobacter species are the most effective microorganisms known for converting organic compounds to electricity in microbial fuel cells and are important in the bioremediation of subsurface environments contaminated with organic and/or metal contaminants. However, the optimal conditions for power production and contaminant bioremediation are significantly different from the low-nutrient, slow growth conditions under which *Geobacter* species are likely to have evolved. Furthermore, genome-scale *in silico* modeling of the metabolism of *Geobacter* species has suggested that *Geobacter* species have metabolic potential far beyond that previously demonstrated in culture. Therefore, it was hypothesized that by exerting the appropriate selective pressure on *Geobacter* species it might possible to increase their effectiveness for practical

applications as well as develop a better understanding of complex physiological processes within these organisms.

For example, although the culture-collection strain of *Geobacter sulfurreducens* produces current at higher power densities than any other microorganism known, the effectiveness of power production was significantly improved with continuous selective pressure for rapid growth on electrode surfaces. In a relatively short time it has been possible to increase overall current production in microbial fuel cells 600%, but even more remarkably, this increased current is produced by far fewer cells. Whereas the culture-collection strain requires a thick biofilm to produce maximal current, similar biofilms are not seen with the evolved strain. Thus, the current production per mg of cell protein is at least 30-fold higher than with the culture-collection strain and continues to improve. The genomes of these evolved strains are currently being sequenced to determine what beneficial mutations have contributed to this remarkable increase in current production capacity.

Another *Geobacter* strain was evolved to transfer electrons to fuel cell anodes at potentials significantly lower than those of the culture-collection strain. This strain has deleted over 6% of its genome. It appears to be adapted not only for electron transfer at low potential, but also for more rapid growth on electrodes and Fe(III) oxide. Potential additional physiological consequences of such a large gene loss are currently under investigation.

More is known about the evolution of another strain of *G. sulfurreducens* that was adapted for rapid extracellular electron transfer to Fe(III) oxide and grows 10-fold faster on Fe(III) oxides than the culture-collection strain. Five mutations were identified in the adapted strain. These were in regulatory genes and genes encoding proteins for fumarate transport and outer membrane protein biogenesis. Microarray analysis comparing gene transcript levels in the adapted strain and the wild type suggested an upregulation in expression of genes encoding: the electrically conductive pili; other extracellular electron transport proteins; and TCA-cycle enzymes. Gene knock-out studies suggested that the pathway for extracellular electron transfer is somewhat different in the evolved strain and biochemical studies demonstrated that the adapted strain has a greater abundance of loosely bound, outer-surface *c*-type cytochromes.

A metabolic feature of *Geobacter* species limiting current production from complex wastes, and restricting the substrates that can be used to promote *in situ* uranium bioremediation, is the limited range of substrates that these organisms can metabolize. For example, the culture-

collection strain of *G. sulfurreducens* is unable to utilize energy/electron-dense compounds, such as sugars and glycerol. Furthermore, although the culture-collection strain of *G. sulfurreducens* can grow on lactate, its doubling time with lactate as the electron donor and fumarate as the electron acceptor is 24 h whereas the doubling time predicted by the *in silico* model for this growth condition is 2.6 h. Repeated transfer of *G. sulfurreducens* in lactate-fumarate medium for over 500 generations reduced the doubling time to 5 h. Furthermore, a lactate-adapted strain was recovered that not only had improved growth on lactate, but also could utilize pyruvate, as well as a number of precursors to pyruvate, such as a variety of sugars and glycerol. The evolution of a strain of *G. sulfurreducens* that can utilize electron-dense fuels, such as sugars and glycerol, that are important components of biomass or wastes (glycerol is a major waste product of biodiesel production) has important implications for current production and bioremediation. Furthermore, it raises an important physiological/ecological question of why *Geobacter* species do not utilize these compounds in soils and sediments, but rather rely on fermentative microorganisms to convert these substrates to acetate and other fermentation products that *Geobacter* species then utilize.

When *G. sulfurreducens* was grown in continuous culture for a year with Fe(III) citrate as the electron acceptor, two of the four strains examined lost the capacity for fumarate reduction whereas strains that had been grown in continuous culture for a year with fumarate as the electron acceptor retained the capacity for Fe(III) reduction. The loss in the capacity for growth on fumarate was associated with mutations in a sigma 54-dependent DNA-binding response regulator that is located immediately upstream of the gene for the fumarate transporter. The loss of a functional fumarate transporter is one of the responses that was also observed in the strain mentioned above which was adapted for rapid growth on Fe(III) oxide. These results suggest that a functional fumarate transporter is deleterious for growth with Fe(III) as the sole electron acceptor. The fact that the fumarate transporter is found in fresh *Geobacter* isolates from a diversity of sedimentary environments suggests that, in nature, a fumarate transporter is a beneficial feature. Further investigation of this hypothesis is underway.

The relative ease in developing strains for enhanced extracellular electron transfer and range of substrate utilization with adaptive evolution contrasts with our previous inability to genetically engineer such large improvements in respiration and metabolism via the enhanced expression of what was considered to be the necessary genes. These results demonstrate that for complex, highly

regulated, poorly understood, physiological properties, adaptive evolution may be the superior initial design tool. Once the mutations leading to the desired modifications are elucidated via adaptive evolution studies, leading to a better understanding of the physiology, then engineering via other strategies may be possible. These studies also suggest that adaptive evolution studies can provide new insights into the physiology and ecology of *Geobacter* species living in subsurface environments.

80 GTL Bioinformatic Analysis of Transcription Regulation of *Geobacter sulfurreducens*

Julia Krushkal¹* (jkrushka@utmem.edu), Jose F. Barbe,¹ Ching Leang,² Toshiyuki Ueki,² Katy Juarez-Lopez,³ Jeanette Peeples,¹ Yanhua Qu,¹ Ronald M. Adkins,¹ and **Derek R. Lovley**² (dlovley@microbio.umass.edu)

¹University of Tennessee Health Science Center, Memphis, Tennessee; ²University of Massachusetts, Amherst, Massachusetts; and ³Instituto de Biotecnología/UNAM, Cuernavaca, México

Project Goals: The overall purpose of this project is to develop experimental and computational tools to predictively model the behavior of complex microbial communities involved in microbial processes of interest to the Department of Energy. The five year goal is to deliver in silico models that can predict the behavior of two microbial communities of direct relevance to Department of Energy interests: 1) the microbial community responsible for in situ bioremediation of uranium in contaminated subsurface environments; and 2) the microbial community capable of harvesting electricity from waste organic matter and renewable biomass. The research in this abstract summarizes research under Subproject IV. The purpose of this subtask is to use computational methods to better understand transcriptional regulation of the expression of environmentally relevant genes in *Geobacter* species.

Geobacter sulfurreducens is a pure culture model representative of the family *Geobacteraceae*, a group of important organisms that participate in bioenergy production and in environmental bioremediation. Our research employs bioinformatic tools to discover transcriptional regulatory mechanisms that are important for *Geobacter* physiology, its response to environmental changes, and its role in bioenergy generation.

Our recent studies have focused on the *G. sulfurreducens* RpoN (σ^{54}) regulon. RpoN is an essential sigma factor, a subunit of RNA polymerase, involved in a variety of cellular processes in *G. sulfurreducens*. Using computational tools, we identified sequence sites in the *G. sulfurreducens* genome that likely represent promoter elements recognized by RpoN. These promoters, predicted using sequence information, were ranked according to their scores, their location, and orientation relative to their target operons, and the change in expression of their target operons in the strain of *G. sulfurreducens* overexpressing the *rpoN* gene as compared to the wild type. We identified RpoN-regulated promoters in the upstream regions of both significantly up- and downregulated operons, which suggests competition by RpoN with other sigma factors. A number of RpoN-regulated promoters with high scores were found upstream of upregulated operons encoding ABC-type branched-chain transporter system proteins and ribosomal proteins, and also upstream of downregulated operons related to flagellar biosynthesis, nitrogen regulation, and ion transport. RpoN-regulated promoters were also identified upstream of operons encoding c-type cytochromes, DNA polymerase subunits, signal transduction components, and other proteins. We further investigated the function of RpoN by analyzing the presence of RpoN-regulated promoter elements near genes with significant changes in expression levels in several deletion mutants of *G. sulfurreducens* (e.g., *omcB* mutant adapted to growth on soluble Fe(III), *pilR* mutant, and others). Several predicted RpoN-regulated promoters have now been experimentally validated. Identification of RpoN-regulated promoters in the genome, combined with functional studies of the RpoN regulon carried out by other members of the *Geobacter* Project, allows a deeper understanding of the importance of RpoN as a global transcriptional regulator of *G. sulfurreducens*.

Corollary to our studies of the RpoN regulon, we are also investigating the PilR regulon. PilR is an enhancer binding protein, which acts cooperatively with RpoN in transcriptional regulation. PilR is involved in the expression of the *pilA* gene, whose product is a structural component of pili, also referred to as nanowires, which are electrically conductive and are required for Fe(III) oxide reduction as well as optimal current production in microbial fuel cells. Using sequence information and microarray analysis of PilR deletion mutant, we used bioinformatic analyses to predict multiple PilR-regulated sites that may affect transcription of operons related to biosynthesis, assembly, and function of pili and flagella, and to cell wall biogenesis. We also documented the co-occurrence of predicted PilR regulatory sites with RpoN-regulated promoters. Experimental validation of these computational predic-

* Presenting author

tions is now underway. This computational discovery of PilR-regulated sites, if confirmed experimentally, along with our knowledge about RpoN-regulated promoter elements will aid in the understanding of the intrinsic nature of molecular mechanisms regulating production and assembly of pili and flagella that are important for Fe(III) reduction and the production of electric current.

We are also continuing our effort to catalogue predicted and reported transcriptional regulatory interactions in *Geobacter*, and we are developing and applying sophisticated software and data mining tools to better analyze and report these interactions. For this purpose, we are continuing to develop and update a database and an accompanying online query system, GSEL (*Geobacter* Sequence Elements), created by our group. This database compiles information on transcription regulatory elements in the genome of *G. sulfurreducens*. We are currently expanding the GSEL database in order to allow its timely update of operon predictions, to incorporate new information on predicted or experimentally validated genome regulatory sites and interactions, and to allow manual curation of database entries, as well as future addition of new search capabilities. The development of the GSEL database and online query system is currently performed using SQL-based database management systems, PHP technologies, and Python frameworks. In addition, we are continuing our ongoing role in the prediction of operon organization of newly sequenced genomes by the *Geobacter* Project. These predictions are being used to better understand transcriptional regulatory mechanisms of gene co-regulation in *Geobacteraceae*. The new information about operon organization is being incorporated in the GSEL database and is also being extensively utilized in experimental studies of *Geobacteraceae*.

81 Computational Modeling of Central Metabolism of *Geobacter* and Related Species

Jun Sun,^{1*} Kai Zhuang,² Karthik Srinivasan,² R. Mahadevan,² Olivia Bui,¹ Tom Fahland,¹ Iman Famili,¹ Mounir Izallalen,³ Carla Risso,³ Shelley Haveman,³ Christophe Schilling,¹ and **Derek Lovley**³ (dlovley@microbio.umass.edu)

¹Genomatica, Inc., San Diego, California; ²University of Toronto, Toronto, Ontario; and ³University of Massachusetts, Amherst, Massachusetts

Project Goals: The goal of this project is to incorporate the wealth of biological data and findings into in silico models, which can describe and predict the likely behavior of *Geobacter* species under various environmental conditions during in situ uranium bioremediation or on the surface of energy-harvesting electrodes. We have implemented an iterative modeling and experimental program to develop a model for central metabolism in *Geobacter sulfurreducens* which accurately predicts the growth and metabolism in this organism under various environmental conditions and has provided a quantitative analysis of factors controlling the competition between Fe(III) reducers in subsurface environments as well as aiding significantly in the functional analysis of the *G. sulfurreducens* genome. In these studies the genomes of various *Geobacteraceae* available in pure culture or assembled from environmental genomic data will be compared on a functional basis by determining the likely metabolic capabilities and responses of these organisms under different environmental conditions with constraint-based in silico modeling. Further, the general methods and approaches developed in these studies will be used to develop genome-based models of other microbial communities that are relevant to processes in the subsurface sediments or on the surface of energy-harvesting electrodes.

Computational modeling on *Geobacter* and related species were conducted in order to better understand the physiology and ecology of *Geobacteraceae* involved in bioremediation and electricity generation from waste organic matter and renewable biomass.

Geobacter metallireducens is of interest because it represents a model for the *Geobacter* species that carry out bioremediation of organic and metal contaminants in subsurface environments. The genome-scale metabolic

* Presenting author

model of *G. metallireducens* was further refined to include 747 genes and 697 reactions. Detailed examination of the refined *G. metallireducens* model suggested that its central metabolism contains several potential ATP-consuming futile cycles, involving energy-inefficient reactions that are not present in the *G. sulfurreducens* model. These *G. metallireducens* unique reactions include the acetyl-CoA synthetase, acetyl-CoA hydrolase, and phosphoenolpyruvate carboxylase reactions. We hypothesized that the energy-inefficient reactions and futile cycles might function to balance key metabolite pools to allow higher fluxes through the TCA cycle for energy production. Experimental biomass yield of *G. metallireducens* growing with pyruvate was lower than optimal *in silico* biomass yield but matched simulation results when fluxes through the futile cycles were assumed. Microarray data of *G. metallireducens* growing with benzoate and acetate indicated that genes coding these unique reactions were up-regulated by benzoate. These results suggested that the futile cycles were likely turned off during *G. metallireducens* growth with acetate for optimal energy utilization, but were up-regulated during growth with complex electron donors to improve flux through TCA cycle for fast energy generation.

Growth of *G. metallireducens* with different electron donors and electron acceptors were simulated using the *in silico* model. Simulation results indicated that: 1) under donor limiting conditions, aromatics compounds such as toluene and cresol allowed the highest biomass yield per substrate and acetate allowed the least; 2) under acceptor limiting conditions, pyruvate produced the highest biomass yield per acceptor; 3) complete oxidization of aromatic compounds required much more electron acceptor than pyruvate; 4) nitrate as electron acceptor resulted in higher biomass yield per substrate or per electron acceptor than Fe(III) or fumarate (in a *G. metallireducens* strain with a dicarboxylic acid transporter to allow the utilization of fumarate). These simulations provided a fast and cost-effective way to understand the metabolism of *G. metallireducens*.

One of the key results from continued genome-scale modeling of *G. sulfurreducens* was the impact of the global proton balance on the biomass yields during reduction of extracellular electron acceptors. In contrast to fumarate reduction, protons generated during acetate oxidation are not consumed during the reduction of Fe(III) as the terminal electron acceptor. The energetic cost of pumping these protons to maintain the membrane potential was implicated in lowered biomass yields during growth with extracellular electron acceptors. We have experimentally evaluated the rate of proton exchange during growth

of *G. sulfurreducens* to confirm the model prediction of proton generation rate.

Furthermore, constraint-based modeling was applied to the investigation of metabolism of *Pelobacter carbinolicus* in the *Geobacteraceae* family. The reconstructed *P. carbinolicus* model contains 740 genes and 705 reactions, and shared 539 reactions with *Geobacter sulfurreducens* metabolic model. The unique reactions of *P. carbinolicus* model reflected some unique metabolic capabilities such as fermentation growth, a second pathway for proline biosynthesis, and the additional reactions in purine biosynthesis. Microarray data were utilized to understand the redundancy associated with some reactions and pathways. To validate the reconstructed model, *P. carbinolicus* model was tested by simulating published growth conditions including fermentations, syntrophic growth, and Fe(III) reduction. Simulation results matched well with experimental data and indicated the accuracy of the current model.

Rhodoferrax ferrireducens is a dissimilatory metal-reducing microorganism which has recently been shown to be abundant in some uranium-contaminated subsurface environments and also has the unique ability to effectively convert sugars to electricity. The reconstructed *R. ferrireducens* genome-scale model contains 737 genes and 756 reactions. The *R. ferrireducens* metabolic model can grow with various electron donors, including acetate, pyruvate, lactate, malate, glucose, benzoate, etc. To understand *R. ferrireducens* growth with these different electron donors, *in silico* simulations were performed with the *R. ferrireducens* model. Under both electron donor and acceptor limiting conditions, glucose allowed the highest biomass yield per electron donor or per electron acceptor. The *R. ferrireducens* modeling research also helped to understand the experimental results that *R. ferrireducens* could not grow with only glucose, but could grow with fumarate serving as both electron donor and acceptor. *R. ferrireducens* appears to be missing a reversible acetaldehyde dehydrogenase that prevents it from growing via glucose fermentation. However, *R. ferrireducens* contains all the genes required to utilize fumarate as electron acceptor allowing *R. ferrireducens* to grow via fumarate fermentation.

In order to better understand the ecology and function of dissimilatory metal-reducing communities in contaminated subsurface environments the dynamics of competition in a metal reducing tri-culture of *R. ferrireducens*, *G. sulfurreducens*, and *E. coli* was modeled. *R. ferrireducens* can oxidize both glucose and acetate while *G. sulfurreducens* is dependent on fermentative microorganisms to produce acetate from glucose. Four sets of two-thousand

* Presenting author

simulations with randomized parameters were performed. The randomized sets are defined by having either high or low maximum glucose feed rate. For the sensitivity analysis, kinetic parameters were varied between 25% and 400% of their literature values. *G. sulfurreducens* dominated in the majority of simulations, confirming the typical observation in subsurface environments. Under glucose-rich conditions, glucose feed concentration shows the highest influence on the *Geobacter* to *Rhodospirillum rubrum* ratio (G/R ratio)—more glucose allows fermenters to exploit the r-strategy more efficiently. However, if the glucose input flux is low, the G/R ratio is mostly dependent on the initial concentration of *G. sulfurreducens*. This result suggests that under these conditions, the fermenters' r-strategy fails and *G. sulfurreducens* success becomes highly dependent on its own initial concentration.

It is surprising that *R. ferrireducens* are outnumbered by *G. sulfurreducens* in anaerobic aquifers and sediments, even though *R. ferrireducens* is metabolically more versatile than *G. sulfurreducens*. Based on simulation results, we propose that the fermenters and *G. sulfurreducens* out-compete *R. ferrireducens* in utilizing glucose and acetate. Under substrate rich conditions, fermenters first out-compete *R. ferrireducens* for glucose using r-strategy, and convert glucose to acetate; then, *G. sulfurreducens* out-competes *R. ferrireducens* for acetate using r-strategy. Under conditions with limited substrate addition, the co-culture lacks sufficient substrate to exploit the r-strategy efficiently, leaving a niche for *R. ferrireducens*.

82 Continued Identification of Small Non-Coding RNAs and Acceptance Rate Studies in Members of the *Geobacteraceae*

Barbara Methé¹* (bmethe@jcv.org), Robert Deboy,¹ Sean Daugherty,¹ Manolito Torralba,¹ Kelly Nevin,² Jonathan Badger,¹ and Derek Lovley²

¹The J. Craig Venter Institute, Rockville, Maryland and

²University of Massachusetts, Amherst, Massachusetts

Project Goals: The *Geobacter* project seeks to apply a systems level biological approach to the study of members of the *Geobacteraceae*. Specific goals of the project presented in this abstract are to apply comparative genomic analyses to understand important aspects of regulation and evolution of these organisms.

Members of the *Geobacteraceae* continue to be the subject of comprehensive studies related to their ability to degrade carbon compounds including many contaminants with the reduction of Fe(III). Further interest in these organisms stems from the practical biotechnological roles that they can play as agents of bioremediation and in energy production. We have begun several analyses that leverage information available from the genome sequences of multiple members of the *Geobacteraceae* to obtain new insights into their evolution and regulation. An update of these analyses is provided here.

In order to better understand the regulatory network of the model organism, *G. sulfurreducens*, as well as other members of the *Geobacteraceae*, we have initiated a study of chromosomally located small non-coding (sRNAs). Increasing evidence suggests that sRNAs exist in numerous organisms where they play important regulatory roles including responses and adaptations to different stresses. We have extended our predictions of sRNA candidates through continued use of the program, sRNAPredict, along with custom written scripts, to compare the *G. sulfurreducens* genome to a total of seven members of the *Geobacteraceae* (*G. metallireducens*, *G. uraniumreducens*, *G. bermidjiensis*, *G. sp. FRC-32*, *G. lovleyi*, *P. carbinolicus*, and *P. propionicus*). The sRNAPredict program utilizes BLAST files generated from intergenic regions (IGRs) of *G. sulfurreducens* against the chromosomes of the other test strains. Positive IGRs were cross-referenced with files generated from TransTerm, RNA motif and QRNA.

From the updated searches, 88 predictions were generated for *G. sulfurreducens*: None of the predictions from *G. sulfurreducens* were found in all of the other seven genomes tested although 17 were found in six of the seven. However, it cannot be ruled out that the distributions of sRNAs could change as three of the test genomes used in these searches are still in a draft phase. The majority of the predictions (34) were found in only one of the other seven genomes. Distributions for the remaining predictions are given as follows: five genomes (4), four genomes (7), three genomes (13), and two genomes (13). Of the 88 predictions, seven have matches to covariance models to non-coding RNAs in the Rfam database. This includes a match to a 6S RNA which is believed to associate with the RNA polymerase holoenzyme containing the sigma70 factor and repress expression from a sigma70-dependent promoter. The presence of this sRNA has been experimentally verified through the use of Northern blot hybridizations and sequence specific primers designed to amplify the sRNA from a population of cDNAs. Although the 6S RNA has been characterized as important in stationary phase growth

—GTL

* Presenting author

we have seen evidence of its expression in both log and stationary phase growth in *G. sulfurreducens*.

An important question yet to be answered comprehensively is whether or not the physical location of a particular family of sRNAs in a query genome is a random event when compared to a family of genomes, or is a “syntenic” relationship maintained. This question has relevant implications in terms of genome evolution and in enhancing the identification of sRNA predictions. An examination of this question was initiated by evaluating whether or not a predicted sRNA in *G. sulfurreducens* and its homolog in a test genome maintain neighboring genes. Eight predictions were analyzed that matched at least three test genomes. Five were found to be associated with a particular gene or set of genes, and each of these sRNA predictions maintained their relative orientation to the neighboring gene(s) in *G. sulfurreducens* and the test genomes. The predictions which “failed” were short sequences with blast e-values above $1e-5$ in some test genomes. These might be pursued by adjusting cutoffs and looking at neighboring genes to confirm correct annotation in the test genomes. These results suggest that sRNAs of a particular family may be associated with a similar gene or gene(s) in a family of genomes.

Evolutionary processes are important forces that drive changes in microbial genomes and they provide a lens through which to better view and understand them. Since advantageous mutations are fixed in a population more rapidly than neutral mutations the rate of nonsynonymous (amino acid replacement) substitution will exceed that of synonymous (silent) substitution if advantageous selection plays a role in the evolution of the protein in question. One method to detect positive selection is to determine if the number of substitutions per nonsynonymous site is significantly greater than the number of substitutions per synonymous site.

To examine sites within predicted proteins that may be undergoing positive selection, Jacquard clusters of predicted proteins encoded in seven genomes of *Geobacteraceae* were created using the SYBIL system for comparative genome analyses. Clusters containing at least five predicted proteins were examined through a pipeline that 1) generated protein alignments 2) created nucleotide alignments of the coding DNA from the protein alignments 3) created a phylogenetic tree based on the sequences in the alignment and 4) ran the PAML program to find sites with significant probability of being under positive selection. Only cases in which the minimum protein length in the alignment was >80% of the maximum were used in order to ensure nearly full length homology. Of the 3,201 clusters considered, 38 had

genes containing sites with evidence of significant positive selection. Classification of the clusters by biological role category revealed that the largest numbers were annotated as conserved hypothetical proteins (8). The other top role categories in terms of membership were transport and binding proteins (6), enzymes of unknown function (5), energy metabolism (4), protein fate (4) and regulatory function (4). An unusual finding was the presence of 200 sites identified in the essential cell division gene, *ftsQ*. Additional analyses of this data is now under way to determine (where protein structures exist) at what locations in the 3D structure of the protein identified sites of positive selection occur and to examine additional data sets for positive selection based on different criteria for cluster membership.

83 GTL Experimental Mapping and Active Annotation of Transcription Initiation Sites of *Geobacter sulfurreducens*

Katy Juárez^{1*} (katy@ibt.unam.mx), Alfredo Mendoza,¹ Leticia Olvera,¹ Maricela Olvera,¹ Enrique Morett,¹ and Derek R. Lovley² (dlovley@microbio.umass.edu)

¹Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, México and ²Department of Microbiology, University of Massachusetts, Amherst, Massachusetts

Project Goals: The main goal of this project is to experimentally map as many transcription start sites (TSS) as possible of genes for which no TSS have yet been described in the literature, using a high throughput technology, and performing a simultaneous active annotation. A database will be generated that will help to better understand promoter structure and gene expression regulation in the *Geobacteraceae*.

Geobacteraceae is the predominant group of bacteria during in situ bioremediation of uranium-contaminated groundwater and on the surface of electrodes harvesting energy from a variety of organic sources. The completion of genome-sequence from multiple members of this group provides the opportunity to apply experimental and computational analyses to obtain new insights into their regulatory processes. Although the availability of the genome sequence of *G. sulfurreducens* has made it possible to predict operon structure, promoters and transcriptional factors binding sites, only a few have been experimentally determined. While such predictions are based mainly on what is known for other bacteria, the

* Presenting author

experimental determination of an important set of transcription start sites will contribute not only to improve the knowledge about the promoter and operon structure in this bacterium, but also increase the predictive capacity for those cases not experimentally determined.

It is well known that the transcription initiation is a critical step in the regulation of gene expression. To know the transcription start sites (TSS) provides relevant information about: a) promoter sequence(s), b) the putative sigma factor(s) and regulatory protein(s) involved in their expression, c) the putative regulatory signals, d) the identity of the initial nucleotide, and also could be used to corroborate operon structure prediction. In order to obtain a global picture of the active promoters in *G. sulfurreducens*, we will follow the large scale TSS mapping methodology that we have previously developed for *E. coli*, which has allowed us to map several hundreds TSS in this organism. To exhaustively map the TSS we will implement two approaches: Directed mapping using a modification of the 5'RACE protocol, and global mapping of TSS using pyrosequencing technology. The modified 5'RACE protocol allowed us to experimentally map more than 300 new TSS in *E. coli*. As controls we mapped the TSS of about 30 genes previously determined experimentally by other groups, confirming the accuracy of our methodology. Interestingly, in this small collection we identified additional TSS for several of these genes not previously detected in the original reports.

We have already tested this methodology in *G. sulfurreducens* and have reported some new TSS, demonstrating that it also efficiently works in this organism. Our strategy for global TSS mapping in *Geobacter* will be to analyze microarray mRNA expression as a guide to select the growth conditions that maximize the expression of as many specific genes and operons as possible. We will start by using the first gene of the predicted operons to map the TSS, we will also analyzed directly the TSS of those genes that have been demonstrated to be physiologically relevant for electron transfer and Fe(III) reduction.

Recently, ultra-high throughput DNA sequencing, using pyrosequencing technology, has allowed rapid determination of the DNA sequence of millions of nucleotides per hour. This technology, implemented in the GS20 instrument of 454 Life Sciences, is based on the sequencing of small individual DNA fragments that are clonally amplified in a very original emulsion format before being sequenced. We have very recently implemented a variation of the standard pyrosequencing methodology to sequence cDNA fragments up to their 3' end. This information will be integrated in a reliable database of

promoters and transcription start sites of *G. sulfurreducens* that will ultimately be linked to analyses of transcriptional regulatory systems conforming a database that can be easily accessed.

84

GTL

Evolution of Electron Transfer Out of the Cell: Comparative Genomics of Nine *Geobacteraceae* Genomes

Jessica E. Butler* (jbutler@microbio.umass.edu), Nelson D. Young,* and Derek R. Lovley (dlovley@microbio.umass.edu)

Department of Microbiology, University of Massachusetts, Amherst, Massachusetts.

Project Goals: See goals for abstract 79.

Geobacter species grow by transferring electrons out of the cell. In the natural environment, this ability presumably arose to exploit abundant but insoluble electron acceptors like Fe(III) oxide. Fortuitously, this same metabolism can be directed to transfer electrons onto energy-harvesting electrodes, or to reduce and remediate subsurface contaminants like uranium.

Previous research on one species, *Geobacter sulfurreducens*, has shown the importance of periplasmic and outer-membrane *c*-type cytochromes in growth by extracellular electron transfer. Previous research on *G. sulfurreducens* has also shown that complete oxidation of fermentation intermediates like acetate via the TCA cycle is the source of these electrons. The work presented here uses the genome sequences of relatives of *G. sulfurreducens* for three related goals: to assess the conservation of previously studied proteins; to identify the other proteins required to form a complete pathway of electron transfer outside the cell; and to model the evolution of this unique metabolic ability.

Completed or 10x-draft genome sequences are available for *G. sulfurreducens*, *G. metallireducens*, *G. uranireducens*, *G. bemidjensis*, *G. strain FRC-32*, *G. lovleyi*, *Pelobacter propionicus*, *P. carbinolicus*, and *Desulfuromonas acetoxidans*. For all of these genomes, gene conservation was predicted using two methods: pairs of orthologous proteins were identified as reciprocal best BLAST matches in an all-against-all comparison of two genomes. Their conservation level was quantified in terms of normalized bit scores, and the selective pressure on them in terms of mutation rates. Orthologous proteins from all genomes were grouped using the Markov cluster algorithm. Genes

* Presenting author

acquired by lateral gene transfer from a source outside the family were also identified. A neighbor-joining phylogeny was inferred for every gene in each genome, and those genes that had either a strongly supported atypical phylogeny or a weakly supported atypical phylogeny and out-of-family best BLAST matches were considered foreign.

There are two different types of metabolism in species related to *G. sulfurreducens*. The *Geobacter* and *Desulfuromonas* species are respiratory – they generate ATP by oxidative phosphorylation and the proton motive force. The *Pelobacter* species are fermentative – they oxidize substrates like acetoin to acetate and generate ATP via substrate-level phosphorylation. We were interested in whether the ability to grow by extracellular electron transfer arose from a fermentative or a mixed-metabolism ancestor. A phylogeny of the family was constructed using proteins that were present in a single copy in each genome. These proteins were concatenated, aligned, and phylogeny was modeled using a Bayesian algorithm. The resulting tree showed that there are two clades of *Geobacteraceae*, with a fermentative *Pelobacter* species in each clade. Since phylogeny alone was insufficient to clarify the ancestral metabolism, we examined the conservation of central metabolism and energy metabolism genes in each of the genomes.

In *G. sulfurreducens*, eight reactions of the TCA cycle oxidize acetate and produce NADH, NADPH, and ferredoxin. An NADH dehydrogenase pump protons across the inner membrane for ATP generation via an ATP synthase. Cytochromes transfer the electrons from the reducing equivalents out to extracellular acceptors. All of the enzymes for acetate transport and oxidation are conserved in all of the completed *Geobacter* genomes, as well as *Desulfuromonas*. One of the NADH dehydrogenases and the ATP synthase were among the very best conserved proteins across the six *Geobacter* species. These enzymes had higher averaged conservation scores than many essential housekeeping genes. Also among the best conserved enzymes in all six *Geobacter* species were several TCA cycle enzymes: the 2-oxoglutarate: ferredoxin oxidoreductase, the succinate dehydrogenase, and the citrate synthase.

The fermentative species have lost several key energy metabolism enzymes. Both *Pelobacter* genomes lack an ortholog for the NADPH-generating step of the TCA cycle, the isocitrate dehydrogenase. In addition, both genomes lack the NADPH dehydrogenase. Both *Pelobacter* genomes have also lost a critical inner membrane enzyme, the cytochrome reductase. All of the genomes of the respiratory species contain at least

one cytochrome *bc* complex that is predicted to transfer electrons from the quinones in the inner membrane out to the periplasmic cytochromes. Neither *Pelobacter* species contains any genes predicted to fill this role, which would leave them unable to transfer electrons to cytochromes. This indicates why these species are unable to directly reduce Fe(III) or to transfer electrons onto an electrode. The genes required for the fermentation pathways in these species include many that are not found in any of the other family members and are most similar to distantly related species. Thus, it appears that the fermentative species evolved from a respiratory ancestor that gained fermentation genes via lateral transfer and lost the ability to transport electrons via cytochromes outside the cell.

The best conserved proteins across all *Geobacter* species included enzymes for acetate oxidation, proton transport, and ATP synthesis. Surprisingly, none of the cytochromes shown to be required for Fe(III) reduction in *G. sulfurreducens* are highly conserved in all the *Geobacter* species. OmcB, a well-studied outer membrane cytochrome, appears to have orthologs in all species based on clustering and genome context, but the sequence conservation is very low, and the number of hemes bound varies. Most of the other cytochromes shown to be important in Fe(III) reduction in *G. sulfurreducens* are not found in all of the *Geobacter* species, including OmcS, OmcT, OmcZ, and MacA. However, the tri-heme periplasmic cytochrome PpcA appears to be important in all of the *Geobacter* species. Orthologs with moderate levels of sequence conservation are found in several copies in all of the genomes. In addition, there are several cytochromes that are polymers of PpcA-like subunits that are found in all of the *Geobacter* species. Thus, most of the key outer membrane cytochromes are not conserved across all species, while the predominant periplasmic cytochrome is. Broader analysis of cytochromes showed that the best conserved cytochromes include many predicted to be localized in or associated with the inner membrane. These include: the four *c*-type cytochromes encoded near the cytochrome *bc* complexes, and a nine-heme cytochrome with seven transmembrane helices, and one subunits of a very well conserved electron transport complex.

Lack of conservation of key cytochromes indicates that one specific pathway of cytochromes is not required for electron transport out of the cell. However, while specific cytochromes may not be conserved, an abundance of cytochromes is a trait of every *Geobacter* genome. All seven of the respiratory species contain at least 100 cytochromes genes. *G. uranireducens* encodes 152 cytochromes, with a total of 1041 hemes, up to 43 per pro-

* Presenting author

tein. Cytochromes appear to be under less selective pressure than other energy metabolism genes. Comparison of *G. sulfurreducens* to *G. bemidjiensis* showed that 95% of the cytoplasmic and inner membrane energy metabolism genes had orthologs, compared to 46% of multi-heme cytochromes. On average, the selection pressure for the conserved cytochromes was 57% of that for the energy metabolism genes. At the extreme, the outer membrane cytochrome OmcS was under 13% of the pressure on the acetate kinase.

These results indicate that the ancestor of this family was respiratory, not fermentative, oxidized acetate, generated ATP by oxidative phosphorylation at the inner membrane, and contained abundant cytochromes for electron transport out of the cell. These results suggest that further study of the cytochrome(s) that provide the essential electric connection between the inner membrane and periplasm is warranted. However, it appears that once protons are moved across the inner membrane allowing ATP synthesis, the path of electrons across the periplasm and outer membrane is likely very variable, and possibly non-specific.

85 GTL Molecular Analysis of the Metabolic State of *Geobacter* Species During *in Situ* Uranium Bioremediation

D.E. Holmes,* P.J. Mouser, H. Elifantz, L.A.L. N'Guessan, M. Akujkar, M.A. Chavan, and D.R. Lovley (dlovley@microbio.umass.edu)

Department of Microbiology, University of Massachusetts, Amherst, Massachusetts

Project Goals: See goals for abstract 79.

Analysis of 16S rRNA sequences, lipids, metagenomic sequences, and environmental proteomics have all unequivocally demonstrated that *Geobacter* species are the predominant organisms catalyzing *in situ* uranium bioremediation at the DOE study site at Rifle, CO. Numerous other studies using one or more of these methods have also suggested that *Geobacter* species are the most abundant dissimilatory metal-reducing microorganisms in a diversity of other contaminated subsurface environments. Therefore, a primary goal of our Genomics:GTL project is to develop genome-based strategies that will make it possible to: 1) elucidate the metabolic state of *Geobacter* species in the subsurface in order to determine factors that might be limiting the bioremediation process and

2) develop genome-scale *in silico* models that can predict how the metabolism and growth of *Geobacter* species will respond to changes in environmental conditions resulting from geochemical heterogeneities and/or environmental changes that are imposed with the goal of optimizing *in situ* bioremediation.

In the last year significant progress has been made in elucidating mechanisms behind the physiological responses of the natural community of *Geobacter* species in the subsurface to important environmental changes. For example, nitrogen is a key nutrient for *Geobacter* species in the subsurface. Our previous analysis of gene transcript abundance in the subsurface during *in situ* uranium bioremediation at the Rifle, CO site demonstrated that *in situ* *Geobacter* species were limited for ammonium and thus had to fix nitrogen from the atmosphere. In order to investigate this phenomenon further, expression of genes for nitrogen fixation and ammonium uptake was evaluated at a series of geochemically diverse wells representing ammonium concentrations ranging from 0-400 μM during the 2007 field experiment at the Rifle site. In one monitoring well in which ammonium concentrations were below detection, the number of *Geobacter* transcripts from *amtB*, which encodes an ammonium transport protein, increased 25 fold when acetate was added to stimulate the growth of *Geobacter* and U(VI) reduction. Transcript levels for *nifD*, a gene involved in nitrogen fixation also increased 27 fold.

Pure culture studies with *Geobacter sulfurreducens* identified *pstB*, which encodes a phosphate transporter, and *phoU*, which encodes a phosphate uptake regulatory protein, as key genes with increased expression during phosphate limitation. Preliminary analysis of gene transcript levels in the subsurface during *in situ* uranium bioremediation suggested that the natural community of *Geobacter* had a significant phosphate requirement as growth was initially stimulated, but that later in the bioremediation process *Geobacter* were probably limited by the availability of other nutrients.

Studies with *G. sulfurreducens* and *G. uraniumreducens*, which was isolated from the Rifle site, identified genes encoding putative acetate transporters that were expressed at high levels when acetate availability limited growth. Transcript levels for these transporters in the natural community of *Geobacter* species were monitored during the course of a field experiment in which the acetate additions were manipulated to alter acetate availability in the groundwater. In accordance with pure culture studies, transcript levels for the acetate transporter genes decreased ca. 20-fold as high (≥ 3 mM) concentrations of acetate became available in the groundwater.

* Presenting author

When acetate additions were temporarily decreased, reducing groundwater concentrations below 1 mM, transcript levels again increased ca. 8-fold.

In order to determine the response of *in situ* *Geobacter* to various nutrient amendments, molecular tools that can be used to monitor changes in the growth rate of *Geobacter* in the environment were developed. Microarray analyses of *G. sulfurreducens* and *G. uraniumreducens* cells grown at different growth rates identified several genes whose expression was directly correlated with cell growth rates. These genes included the ribosomal proteins, *rplL* and *rpsC*, and the cell division proteins, *ftsA* and *ftsZ*. Expression of these genes by *G. uraniumreducens* was then monitored over time during growth in sediments and on the soluble electron acceptor, fumarate, at 18°C, 30°C, and 37°C. Growth of *G. uraniumreducens* in sediments was significantly slower at higher temperatures, and the number of mRNA transcripts from all 4 of these genes at mid log phase was also lower at the higher temperatures. In addition, the number of mRNA transcripts from *ftsZ* and *rpsC* corresponded with the number of cells and acetate concentrations in the fumarate-grown cultures. Similar results were observed in groundwater collected during the uranium bioremediation field experiment.

In order to refine the existing *Geobacter* genome-based *in silico* model, the genes that are found in the *Geobacter* species that predominate in subsurface environments were further investigated. This was accomplished via: 1) isolation of several new *Geobacter* strains representative of the *Geobacter* species that predominate in the subsurface; 2) direct sequencing of genomic DNA collected during *in situ* uranium bioremediation; and 3) amplifying and sequencing genomic DNA from single cells isolated from the subsurface. Phylogenetic analysis of this genome data indicated that *Geobacter* species at these sites typically cluster into a tight phylogenetic group, referred to as Subsurface Clade 1. Several subsurface clade 1 genomes have been completed or are in the process of being sequenced. Analysis of these genomes has indicated that they all have complete TCA cycles and numerous c-type cytochrome genes. Analysis of the complete *G. uraniumreducens* genome showed that it is significantly larger than those of other *Geobacteraceae*, and is marked by an abundance of mobile genetic elements (~3% of genes) and other features indicative of active gene exchange, duplication, rearrangement and loss. Large insertion/deletion events in flagellar structural genes could explain the long flagella of *G. uraniumreducens*.

The genomes of *G. bemidjiensis* and strain M21 are particularly similar, even though they were recovered from two geographically and geochemically distinct sub-

surface environments. *G. bemidjiensis* was isolated from a petroleum contaminated aquifer located in Bemidji, MN, while strain M21 was isolated from the uranium-contaminated aquifer in Rifle, CO. Comparison of *G. bemidjiensis* and strain M21 genomes revealed that more than half of the c-type cytochromes present in *G. bemidjiensis* have homologues in strain M21. In addition, unlike other *Geobacter* species, the genomes of *G. bemidjiensis* and strain M21 contain *E. coli*-like glucokinase, galactokinase, galactose-1-phosphate uridylyltransferase, and hydroxypyruvate reductase genes which are all involved in glucose and galactose catabolism. In addition, the glucose/galactose transporter was detected in *G. bemidjiensis*. Preliminary analysis of metagenomic and single cell data showed that the majority of sequences in these libraries are most similar to *G. bemidjiensis* and strain M21.

86 GTL Molecular Mechanisms Regulating Gene Expression in *Geobacter* *sulfurreducens* Under Environmentally Relevant Conditions

Toshiyuki Ueki^{1*} (tueki@microbio.umass.edu), Ching Leang,¹ Maddalena V. Coppi,¹ Regina A. O'Neil,¹ Kelly P. Nevin,¹ Trevor L. Woodard,¹ Anna Liu,² Barbara A. Methé,³ and **Derek R. Lovley**¹ (dlovley@microbio.umass.edu)

¹Department of Microbiology and ²Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts and ³J. Craig Venter Institute, Rockville, Maryland

Project Goals: See goals for abstract 79.

In order to predictively model the activities of *Geobacter* species during *in situ* uranium bioremediation and on the surface of electrodes harvesting electricity from waste organic matter and renewable biomass, we have been investigating the molecular mechanisms regulating gene expression under various environmental conditions in a representative of *Geobacter* species, *Geobacter sulfurreducens*. This information is essential for improved *in silico* models that are being employed in the optimization of current production and bioremediation practices.

The number of homologs of regulatory proteins such as transcription factors and signal transduction proteins is large in order for organisms to promptly and properly sense and respond to a variety of environmental changes.

* Presenting author

Therefore, it is important to understand the function of each homolog. It is, however, necessary to develop a system-level analysis for the elucidation of regulatory networks in global gene expression. Such an analysis can be integrated by functional (e.g. microarray analysis) and comparative genomics. We are developing a systematic analysis to elucidate gene regulation in *Geobacter* species, as microarray analyses under a variety of conditions and genome sequences of many *Geobacter* species are available. For example, we recently applied a systematic analysis integrated by functional and comparative genomics in combination with biochemical and genetic methods to elucidate gene regulation during nitrogen fixation. This is of interest because our previous studies have demonstrated that *Geobacter* species are forced to utilize atmospheric nitrogen as a nitrogen source during *in situ* uranium bioremediation. We identified novel regulatory cascades controlling gene expression during nitrogen fixation in *G. sulfurreducens*. The cascades consisted of the two two-component systems GnfL/GnfM and GnfK/GnfR. The GnfL/GnfM system appeared to be the master regulator for the expression of genes involved in nitrogen fixation. It is likely that the GnfK/GnfR system was under the control of the GnfL/GnfM system and regulated a subset of genes by antitermination. Mutants of GnfK and GnfR were able to grow in the presence of fixed nitrogen, while they were unable to grow in the absence of fixed nitrogen. Such a systematic analysis integrated by functional and comparative genomics in combination of biochemical and genetic methods should also be applicable to the elucidation of other gene regulation in *G. sulfurreducens* as well as other *Geobacter* species.

The two-component system is an important strategy for sensing and responding to environmental conditions. It typically consists of a sensor kinase, which senses an environmental signal, and a response regulator, which regulates gene expression to adapt to an environmental change. The genome of *G. sulfurreducens* encodes an unusually large number of proteins belonging to the two-component system, which may reflect the need to adapt to a myriad of different conditions in subsurface environments. The two-component system GsuTCS1 was found to be involved in Fe(III) reduction as well as redox sensing. The sensor kinase of GsuTCS1 has a unique sensor domain containing *c*-type heme binding motifs. To further understand the function of GsuTCS1, a strain over-expressing the sensor kinase and the response regulator of GsuTCS1 was constructed. This strain showed biofilm formation under a condition where the wild-type strain does not form biofilms, suggesting that GsuTCS1 regulates genes involved in biofilm formation. Biofilm formation has been shown to be an important cellular

process for optimal current production in microbial fuel cells.

Our previous studies have demonstrated that *in situ* transcript levels of the gene for citrate synthase in groundwater during *in situ* uranium bioremediation can be used to infer rates of *Geobacter* metabolism in the subsurface. Citrate synthase is a key enzyme in the TCA cycle and regulates the entry of carbon into the TCA cycle. Investigations into the mechanisms regulating expression of *gltA*, the citrate synthase gene, demonstrated that the primary transcriptional regulator was a repressor. Genome-wide sequence analysis identified additional genes, whose putative promoters contain a DNA element with high similarity to the repressor binding site in the *gltA* promoter and which are likely involved in biosynthesis and energy generation, in *G. sulfurreducens* as well as other *Geobacter* species. This suggests that the transcriptional repressor involved in *gltA* expression is a global regulator for biosynthesis and energy generation among *Geobacter* species. By a systematic analysis combined with comparative genomics and a biochemical assay, several transcription factors were identified as candidates for the repressor. These insights into the regulation of central metabolisms provide important information for modeling the cellular activities of *Geobacter* species.

G. sulfurreducens has previously been considered a strict anaerobe. However, it has been shown that *G. sulfurreducens* is able to tolerate exposure to atmospheric oxygen and to grow with oxygen as the sole electron acceptor. These results may explain how *Geobacter* species rapidly become the predominant Fe(III)-reducing microorganisms in a diversity of subsurface sediments, when organic electron donors are introduced into previously oxic subsurface sediments in order to stimulate anaerobic dissimilatory metal reduction. Two genes encoding putative terminal oxidases, cytochrome *c* oxidase and cytochrome *bd* ubiquinone oxidase, were identified in the *G. sulfurreducens* genome. Cytochrome *bd* ubiquinone oxidase was found to play a key role in oxygen metabolism. The expression of the gene encoding cytochrome *bd* ubiquinone oxidase was under the control of multiple factors such as the stationary-phase sigma factor RpoS, FNR homologs and PerR homologs. These findings add significant knowledge to our understanding of the molecular strategies, which *Geobacter* species utilize to survive oxic environments that they encounter in the subsurface.

One of the defining features of *Geobacter* species is a three gene cluster consisting of two genes encoding Fe(II)-dependent transcriptional repressors, *fur* and *ideR*, flanking a gene for a ferrous uptake protein, *feoB-1*. Because *Geobacter* species can be exposed to a wide range

of Fe(II) concentrations in the subsurface and have unusually high assimilatory iron requirements due to their exceptionally high *c*-type cytochrome content, tight control over intracellular iron homeostasis is likely to be critical for their survival. Phenotypic and microarray analyses with *G. sulfurreducens* mutants lacking *fur* and *ideR* indicated that both Fur and IdeR function as Fe(II)-dependent transcriptional repressors that simultaneously regulate genes involved in intracellular iron homeostasis, protein folding and metabolisms. Thus, these genes may play a critical role in allowing *Geobacter* species to respond to and thrive in fluctuating Fe(II) concentrations in the subsurface.

Our understanding of regulatory mechanisms at the molecular levels in *Geobacter* species has been expanding in quantity as well as quality. These studies allow us to improve our models for predicting how *Geobacter* species respond to various changes in environmental conditions during *in situ* bioremediation and energy harvesting.

87

Towards a Transcription Regulatory Network (TRN) in *Geobacter sulfurreducens*

Karsten Zengler^{1*} (kzengler@bioeng.ucsd.edu), Yu Qiu,¹ Byung-Kwan Cho,¹ Larisa Kagan,¹ James Elkins,¹ **Derek R. Lovley**,² and Bernhard Palsson¹

¹University of California, San Diego, California and
²University of Massachusetts, Amherst, Massachusetts

A comprehensive study of genome-wide protein-DNA interactions was performed with *E. coli* as a model organism. Chromatin immunoprecipitation (ChIP) coupled with high-density tiling arrays (ChIP-chip) has been applied to perform genome-wide location analysis for three classes of DNA proteins in *E. coli*. Proteins examined were RNA polymerase and sigma factors, broad acting transcription factors as well as DNA bending proteins. The genome-scale transcriptional regulatory network established for *E. coli* allows for the first time in depth understanding of fundamental aspects of chromosome structure, DNA replication, DNA repair, response to stress and regulation of metabolism.

Towards a comprehensive understanding of the transcriptional regulatory network in *Geobacter sulfurreducens* we adapted the protocol established in our laboratory for *E. coli* to work in *G. sulfurreducens*. *G. sulfurreducens* is capable of transferring electrons to a variety of electron

* Presenting author

acceptors including Fe(III), U(IV), and the surface of electrodes making it the candidate of choice for bioremediation of contaminated environments and harvesting electricity from waste organic matter. In depth understanding of how *G. sulfurreducens* functions will have a great impact on optimizing bioremediation and energy harvesting applications. Genome-wide binding patterns of RNA polymerase and σ^{70} in *G. sulfurreducens* were determined by ChIP-chip analysis. Patterns of RNA polymerase and σ^{70} binding were compared from cells grown with acetate under fumarate- and Fe(III)-reducing conditions. All ChIP-chip results were complemented by gene expression profiles using high-density tiling arrays. In addition promoter regions in *G. sulfurreducens* were determined under various growth conditions by ChIP-chip analysis of cells treated with rifampicin (rifampicin inhibits transcription elongation by RNA polymerase but not its binding to promoter region). **Polyclonal antibodies** against further sigma factors from *G. sulfurreducens* (RpoE, RpoH, RpoN, and RpoS) are currently produced. In addition, we established a system for routine epitope-tagging of transcription factors in *G. sulfurreducens*.

GTL

88

Structural Characterization of *Geobacter sulfurreducens* Heme Proteins: Two Novel Periplasmic Sensor Domains from Chemotaxis Proteins, and the Soluble Part of a Membrane Protein OmcF

P. Raj Pokkuluri,¹ Yuri Y. Londer,¹ Norma Duke,¹ Stephan Wood,¹ Miguel Pessanha,² Leonor Morgado,² Teresa Catarino,³ Carlos A. Salgueiro,² and **Marianne Schiffer**^{1*} (mschiffer@anl.gov)

¹Biosciences Division, Argonne National Laboratory, Argonne, Illinois; ²Requimte, Centro de Química Fina e Biotecnologia, Dep. Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal; and ³Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal

Project Goals: As sub-project of GTL grant "Genome-based models to optimize *in situ* bioremediation of uranium and harvesting electrical energy from waste organic matter, Derek Lovley (PI)" our goals are to analyze selected proteins to understand their function in the cell. This includes modeling of structures based on their amino acid sequences, determination of their

GTL

structures, and the functional interpretation of the structures, such as active sites and surface properties.

Geobacter sulfurreducens encodes over 100 cytochromes containing *c*-type hemes. We determined the structures of periplasmic sensor domains of two methyl-accepting chemotaxis proteins encoded by GSU0582 and GSU0935 (1) and the structure of the soluble part of a membrane protein OmcF encoded by GSU2432 (2).

The heme containing sensor domains (about 135 residues) were expressed in *E. coli*, characterized in solution, and their crystal structures were determined. The R factor is 24.7% for 2.0 Å data and 20.5% for 1.9 Å for GSU0582 and GSU0935 sensors, respectively. In the crystal, both sensor domains form “swapped” dimers and reveal a novel way of forming PAS-type domains using two chains. These are the first PAS domain structures that contain a covalently bound heme. The swapped segment consists of two helices of about 45 residues at the N-terminus with the hemes located between the two monomers. In the case of GSU0582 sensor, the dimer is related by a crystallographic 2-fold axis and the heme is coordinated by an axial His and a water molecule. In the case of GSU0935 sensor, the crystals contain a non-crystallographic dimer, and surprisingly, the coordination of the heme in each monomer is different; Monomer A heme has His-Met ligation and Monomer B heme has His-water ligation as found in the GSU0582 sensor. Optical absorption, EPR and NMR spectroscopies have revealed that the heme groups of both sensor domains are high- and low-spin in the oxidized and reduced forms, respectively, and that the spin-state interconversion involves a heme axial ligand replacement. The reduction potentials of the sensor domains of GSU0582 and GSU0935 are -156 mV and -251 mV, respectively.

At present, we do not know what compound or compounds these chemotaxis proteins sense in *G. sulfurreducens* but homologous proteins are conserved in the *Geobacteraceae* family. Even though the two sensors have similar structures and spectroscopic properties, the reduction potentials of the heme groups are quite distinct, differing by approximately 100mV, suggesting that they might perform their physiological functions in environments with different redox potentials. We propose that these periplasmic sensor domains could be part of a global cellular mechanism for sensing the periplasmic redox potential or small ligands. The swapped dimerization of these sensor domains and redox-linked ligand switch might be related to the mechanism of signal transduction by these chemotaxis proteins. We suggest that the helix swapped dimer formation could be a mechanism for signal transduction from the periplasm to

the cytoplasm as the swapped dimer will alter the relative positions of the trans-membrane helices that can lead to a change in the effector part of the molecule located in the cytoplasm.

OmcF encoded by GSU2432, was identified as an outer-membrane heme protein that affects the expression of the membrane proteins, OmcB and OmcC, in *G. sulfurreducens* (2). The OmcF deficient strain is impaired in its ability to grow on Fe(III) citrate (2). OmcF is a 104 residue protein with one heme attachment site and a lipid anchor. The soluble part of OmcF is homologous to cytochromes *c₆* found in algae and cyanobacteria, where the cytochrome *c₆* transfers electrons from cytochrome *b_f* to photo system I. The soluble part of OmcF, from residue 20, was expressed in *E. coli* and its structure was determined by X-ray diffraction; the R-factor is 17.0% for 1.35 Å data. It is a protein with mostly helical structure. The heme is His - Met coordinated. The *c₆* from green algae *Monoraphidium braunii* (3) is a very close structural homolog with a DALI score, Z=13.1. The reduction potential of OmcF is about +129 mV; interestingly the midpoint potential of *M. braunii c₆* is +358 mV (4).

X-ray diffraction data were collected at Structural Biology Center beam line of the APS.

References

1. Londer YY, Dementieva IS, D'Ausilio CA, Pokkururi PR and Schiffer M (2006). Characterization of a *c* type heme containing PAS sensor domain from *Geobacter sulfurreducens* representing a novel family of periplasmic sensors in *Geobacteraceae* and other bacteria. *FEMS Microbiol Lett* 258: 173-181.
2. Kim B-C, Leang C, Ding Y-HR, Glaven RH, Coppi MV and Lovley DR (2005). OmcF, a Putative *c*-Type Monoheme Outer Membrane Cytochrome Required for the Expression of Other Outer Membrane Cytochromes in *Geobacter sulfurreducens*. *J. Bacter.* 187: 4505-4513.
3. Frazo C, Soares CM, Carrondo MA, Pohl E, Dauter Z, Wilson KS, Hervais M, Navarro JA, De la Rosa MA and Sheldrick GM (1995). Ab initio determination of the crystal structure of cytochrome *c₆* and comparison with plastocyanin. *Structure*, 3: 1159-1169.
4. Campos AP, Aguiar AP, Hervas M, Regalla M, Navarro JA, Ortega JM, Xavier AV, De La Rosa MA and Teixeira M (1993). Cytochrome *c₆* from *Monoraphidium braunii*. A cytochrome with an unusual heme axial coordination. *Eur. J. Biochem.* 216: 329-341.

89

Dynamic Systems-Level Analysis of Oxygen-Dependent Cell State Transitions Reveals a Surprising Chronology of Cellular Events

Amy K. Schmid,¹ David J. Reiss,¹ Amardeep Kaur,¹ Min Pan,¹ Nichole King,¹ Phu T. Van,¹ Laura Hohmann,¹ Daniel B. Martin,^{1,2} and **Nitin S. Baliga**^{1*} (nbaliga@systemsbiology.org)

¹Institute for Systems Biology, Seattle, Washington and ²Divisions of Human Biology and Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington

Adjustment of physiology in response to changes in oxygen availability is critical for the survival of all organisms. However, the chronology of events and the regulatory processes that determine how and when changes in environmental oxygen tension result in an appropriate cellular response is not well understood at a systems level. Therefore, transcriptome, proteome, ATP, and growth changes were analyzed in a halophilic archaeon to generate a temporal model that describes the cellular events that drive the transition between the organism's two opposing cell states of anoxic quiescence and aerobic growth. According to this model, upon oxygen influx, an initial burst of protein synthesis precedes ATP and transcription induction, rapidly driving the cell out of anoxic quiescence, culminating in the resumption of growth. This model also suggests that quiescent cells appear to remain actively poised for energy production from a variety of different sources. Dynamic temporal analysis of relationships between transcription and translation of key genes suggests several important mechanisms for cellular sustenance under anoxia as well as specific instances of posttranscriptional regulation. Importantly, genes which code for protein complexes involved in energy generation appear to be regulated in tandem. For instance, aerobic metabolism complexes such as the ribosome, ATP synthase, and succinate dehydrogenase appear to have similar time lags between transcription and translation. Time lagging analysis of anaerobic energy generation complexes is currently underway.

GTL

90

Rhythmic Gene Expression in *Halobacterium* NRC-1 After Day/Night Cycle Entrainment

Kenia Whitehead,¹ Min Pan,¹ Kenichi Masumura,² Tetsuya Mori,³ Lena Suzuki,³ Carl. H. Johnson,³ Richard Bonneau,⁴ and **Nitin S. Baliga**^{1*} (nbaliga@systemsbiology.org)

¹Institute for Systems Biology, Seattle, Washington;

²Division of Genetics and Mutagenesis, National Institute of Health Sciences, Tokyo, Japan;

³Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee; and ⁴Center for

Genomics and Systems Biology, New York University, New York, New York

Small protein complexes forming biological oscillators are used by many organisms to coordinate gene expression, metabolism and cell physiology. Thus far, molecular oscillators for day/night cycles have not been identified in prokaryotes with the sole exception of cyanobacteria. Our goal is to study day/night rhythms in gene expression in the halophilic archaeon, *Halobacterium* NRC-1 during and after entrainment. Recent results have shown that ~10% of the genome follows a pattern mediated by light/dark cycles with peaks in gene expression at the day/night transition even after the day/night stimulus has been removed and cells are in constant darkness. Rhythmically expressed genes encode enzymes at key positions in carbon and nitrogen metabolism as well as a host of transcriptional regulators. While the molecular basis for this rhythm generation and memory in *Halobacterium* NRC-1 is unknown we can now begin to examine the protein complexes mediating rhythm generation in archaea.

GTL

* Presenting author

91

Transcriptome Structure of *Halobacterium* sp. NRC-1 and Its Dynamic Changes During Growth

Tie Koide,¹ Marc Facciotti,¹ David Reiss,¹ Kenichi Masumura,² Christopher Bare,¹ Min Pan,¹ and **Nitin Baliga**^{1*} (nbaliga@systemsbiology.org)

¹Institute for Systems Biology, Seattle, Washington and ²Division of Genetics and Mutagenesis, National Institute of Health Sciences, Tokyo, Japan

Project Goals: MAGGIE integrates teams at Lawrence Berkeley National Lab and the Advanced Light Source (ALS) with researchers at the Scripps Research Institute, the University of Georgia, the University of California Berkeley, and the Institute for Systems Biology to achieve a molecular-level understanding of the dynamic macromolecular machines that underlie all of microbial cell biology. MAGGIE is providing improved technologies and comprehensive characterizations to efficiently couple gene sequences and genomic analyses with protein interactions and thereby elucidate functional relationships and pathways. The operational principle guiding MAGGIE objectives can be succinctly stated: protein functional relationships involve interaction mosaics that self-assemble from independent protein pieces that are tuned by modifications and metabolites.

Microarray analysis of transcriptome changes in *Halobacterium salinarum* sp. NRC-1 have revealed that nearly half of all annotated genes undergo some significant change in expression to help mediate the physiological transitions that occur during growth. In most cases, dynamic assemblies of molecular protein complexes at key genomic loci regulate these changes. In order to understand how these transcriptional complexes contribute to the regulation of gene expression we must first understand precisely when and where they assemble in the genome. One potentially key indicator for localizing complex assembly is to map the transcript boundaries throughout the genome. Using high-resolution tiling arrays, we have constructed a transcriptome structure map generated for *Halobacterium* sp. NRC-1 from cells harvested at various optical densities throughout a standard laboratory growth experiment. This has enabled us to accurately map transcript boundaries including features such as alternative transcription start sites and operon boundaries. When integrated with existing ChIP-chip (protein-DNA interaction) data the detailed and dynamic transcriptome map allows for great insight

GTL

into the assembly of transcriptional complexes and their influence on patterns of global gene expression. Furthermore, we have also revealed the expression of nearly 150 putative non-coding regions, nearly half of which are differentially transcribed during growth. These non-coding RNAs are poorly understood in archaeal systems but have been shown to act as important post-transcriptional regulators in bacteria and eukaryotes by altering the stability of mRNAs, translation initiation and the assembly of protein complexes. The interplay between these putative regulatory RNAs and their targets is currently under investigation.

92

The *Caulobacter* Divisome: Parts List, Assembly, and Mechanism of Action

Erin D. Goley,¹ Yi-Chun Yeh,¹ Luis Comolli,² J.N. Werner,¹ Ken Downing,² Zemer Gitai,³ **Harley H. McAdams**¹ (hmcadams@stanford.edu), and Lucy Shapiro^{1*}

¹Stanford University School of Medicine, Stanford, California; ²Lawrence Berkeley National Laboratory, Berkeley, California; and ³Princeton University, Princeton, New Jersey

Project Goals: The objective of the project is development of a global and integrated view of the dynamically changing structures involved in the bacterial cell's function, cell cycle progression, response to environmental challenges, and polar development. We seek to identify multiprotein complexes associated with regulation of bacterial development, chromosome replication and segregation, cytokinesis, and intracellular uranium sequestration, as well as the regulatory processes and protein complexes that construct, position, and degrade these structures.

The cell division apparatus (divisome) in bacteria mediates the constriction of the cell membranes and the inward growth of the cell wall in coordination with cell growth and chromosome segregation. Despite decades of study, however, relatively little is known about the structure and assembly of the divisome or the molecular functions of its components. The tubulin relative FtsZ is the best characterized and most highly conserved divisome protein. It is a GTPase that polymerizes near midcell, defining the site of cell division. FtsZ serves as a scaffold for assembly of the divisome and is hypothesized to generate constrictive forces. As the structure, dynamics and function of FtsZ are likely to be regulated by inter-

GTL

* Presenting author

acting partners, we sought to identify all FtsZ-binding proteins in *Caulobacter*. To do this, we developed an assay in which we overproduce a GTPase-defective mutant of FtsZ, causing cells to adopt a distinct morphology wherein long, slender constrictions containing FtsZ separate the cell bodies. We found that fluorescent fusions to *Caulobacter* homologs of all known FtsZ-binding proteins co-localize with FtsZ in the constrictions of mutant cells, whereas divisome proteins that do not bind to FtsZ are diffuse. To identify novel FtsZ-binding proteins we used this assay to probe the library of *Caulobacter* strains bearing fluorescent fusions to 442 different localized proteins generated by the Gitai lab at Princeton. From this screen, we identified six proteins that clearly localized to the constrictions, four of which are previously uncharacterized. To date, we have confirmed that three of these proteins bind directly to FtsZ *in vitro*. Surprisingly, we discovered an additional set of proteins in our screen that were specifically excluded from the FtsZ-rich constrictions, indicating a second mode of FtsZ-directed protein localization in *Caulobacter*. We are now taking genetic, cytological, and biochemical approaches aimed at uncovering the mechanisms by which FtsZ controls the localization of these factors and determining their cellular functions.

In addition to those new FtsZ-binding proteins identified in our screen, at least sixteen other proteins are recruited to form the divisome downstream of FtsZ. To gain molecular insight into their functions, we have cloned the *Caulobacter* homologs of the known cell division proteins and have so far confirmed that thirteen of these are essential for cytokinesis in this organism. Moreover, we find that fluorescent fusions to each of these proteins localize to the division site and we are in the midst of a careful study following the timing of their localization to better understand the process of divisome assembly. Finally, we have made significant progress in understanding the role of the Tol-Pal complex in *Caulobacter* cytokinesis. Using a combination of genetic, cell biological, and high resolution microscopy techniques (in collaboration with the Downing lab at LBNL) we have shown that Pal is specifically required for outer membrane integrity and for invagination of the outer membrane during cytokinesis. Current and future efforts are directed at similarly detailed structural and functional analyses of the other *Caulobacter* divisome proteins.

This work was funded by the DOE GTL project titled "Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis." The PI of this multi-institutional grant is Harley McAdams, Stanford University.

* Presenting author

93 Development of Methods for Correlated Light and Electron Microscopic Analysis of Protein Complexes: Application to *Caulobacter crescentus*

Guido M. Gaietta,¹ Sean Prendergast,¹ Ying Jones,¹ James Obayashi,¹ Thomas J. Deerinck,¹ Grant Bowman,² Yi Chun Yeh,² Lucy Shapiro,² **Harley McAdams**² (hmcadams@stanford.edu), and Mark H. Ellisman^{1*}

¹National Center for Microscopy and Imaging Research, Department of Neurosciences, University of California, San Diego, California; ²Department of Developmental Biology, Stanford University, School of Medicine, Beckman Center, Stanford, California; and ³Life Science Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California

Project Goals: The specific aims in the original proposal were: (1) Adapt 4Cys tags and related compounds to the labeling and analysis of protein complexes in *Caulobacter crescentus*. (2) Enhance the throughput of the end-to-end electron tomographic data collection and volume generation process.

Microbial cells present interesting challenges to multi-scale microscopic analysis by correlated light and electron microscopic imaging (LM and EM). Challenges include limited permeability of the cell wall, sensitivity of the cells to conventional chemical fixatives and their small size. We have systematically addressed these obstacles and developed methods that provide for high quality ultrastructure and enable correlation of fluorescent signals with electron density of corresponding subcellular structures in 3D electron tomographic reconstructions. These methods are now being extended to allow detection of signals from specific complexes of interest in *Caulobacter crescentus*, as originally delineated in the GTL project *Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis*. Specifically:

1. To optimize preservation of supramolecular structure in fluorescently photoconverted specimens we have developed methods that combine fluorescence photoconversion with high-pressure freezing and freeze-substitution (HPF/FS). This allows use of rapid primary aldehyde fixation and labeling of specific protein complexes by fluorescence photocon-

version with the advantages of HPF, which avoid the bulk specimen shrinkage due to solvent dehydration. Samples prepared for electron tomography and 3D reconstruction using this new approach exhibit spectacular ultrastructure and overall form.

- To further enhance the detection of lower abundance proteins we are exploring approaches that increase the deposition of electron dense products at sites bearing fluorescent reporters. New methods include the incorporation of multiple tetracysteine motifs to the target protein/fluorescent protein complex and enhancing rates of photoconversion by utilizing more reactive derivatives of the diaminobenzidine chromogen as well as the addition of catalytic cofactors.
- Recent developments in super-resolution LM have stimulated us to devise alternative approaches that overcome the aforementioned obstacles. These new approaches have been designed to be compatible with methods like photoactivation localization microscopy (PALM), stochastic optical reconstruction microscopy (STORM) or stimulated emission depletion (STED) microscopy, and enable the acquisition of images containing information well beyond the diffraction limit of conventional light optical systems. The key enabling technology is the development of procedures that maintain fluorescent signals in fixed and resin embedded specimens prepared by HPF and epoxy fixation/FS. The ability to simultaneously preserve fluorescent signals and cellular ultrastructure after embedding in epoxy resins represents a powerful alternative to fluorescent photoconversion and should greatly enhance the ability to detect lower abundance proteins. The applicability of this approach is further increased by using fusions with photoactivatable/photoswitchable fluorescent proteins, which retain their function after embedding and allow combined high super-resolution LM protein localization followed by directly correlated electron microscopy and 3D reconstruction using electron tomography.

This work was funded by the DOE GTL project titled "Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis". The PI of this multiinstitutional grant is Harley McAdams, Stanford University.

94 Mapping Protein-Protein Interaction Networks of *Caulobacter crescentus* using Tandem Affinity Purification

Jian Zhu,¹ Ping Hu,¹ Tom Taverner,² Gary Andersen,¹ Lucy Shapiro,² **Harley McAdams**² (hmacadams@stanford.edu), and Thomas Earnest^{1*} (TNEarnest@lbl.gov)

¹Lawrence Berkeley National Laboratory, Berkeley, California and ²Stanford University, Palo Alto, California

Project Goals: Dynamic localization of multiprotein complexes and their structural studies at molecular and cellular levels.

Tandem Affinity Purification (TAP) has proven to be a powerful tool for the investigation of protein-protein interactions in yeast and in *E. coli* as it can purify stable multi-protein complexes in near native conditions. We have developed methods of integrating the TAP tag with the target ORF at the C-terminal using the Gateway system to allow for higher target throughput to be achieved. Tagging of the N-terminal can be accomplished by a more complicated protocol for targets of high scientific value where C-terminal tagging is not possible, when the tag would interfere with the C-terminal interactions with other proteins required for proper function, or when the C-terminal is cleaved during proteolytic regulation. In our initial test set, which consisted of targets covering a wide range of sizes and functionality, approximately 75% of the targets could be C-terminal TAP tagged and purified through the second affinity column. Investigations of ORFs targeted due to their importance in cell polarity, cell division, and cell cycle control also exhibited a high success rate for tag incorporation and purification. For example the hybrid histidine kinase, CckA, co-purifies with a number of proteins not pulled down from wild-type cells after identical growth, disruption, and purification. Mass spectroscopic analysis of the other proteins that co-purify is underway. Components of multiprotein complexes that are identified by this approach are then over expressed by a variety of methods to produce multi-protein complexes and sub complexes in sufficient quantities for structural studies to be performed.

Funding from the DOE Genomics:GTL project titled "Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis." The Principal Investigator of this multi-institutional grant is Harley McAdams, Stanford University.

* Presenting author

95

GTL

A Polymeric Protein Anchors the Chromosome Origin and Governs Multiprotein Complex Assembly at a Bacterial Cell Pole

Grant Bowman,¹ Luis Comolli,² Marcelle Koenig,¹ Jian Zhu,² Thomas Earnest,² Ken Downing,² W. E. Moerner,¹ and **Lucy Shapiro**^{1*} (shapiro@stanford.edu)

¹Stanford University, Sanford, California and
²Lawrence Berkeley National Laboratory, Berkeley, California

Project Goals: The objective of the project is development of a global and integrated view of the dynamically changing structures involved in the bacterial cell's function, cell cycle progression, response to environmental challenges, and polar development. We seek to identify multiprotein complexes associated with regulation of bacterial development, chromosome replication and segregation, cytokinesis, and intracellular uranium sequestration, as well as the regulatory processes and protein complexes that construct, position, and degrade these structures.

In rod shaped bacteria, the cell poles are often utilized as positional reference points for the localization of cellular structures, such as flagella and stalks, polar regulatory complexes of phospho-signaling proteins, and protein complexes that regulate cell division. In some species, the chromosomal origins are also targeted to the cell poles. We find that multiple aspects of cellular organization are linked in *Caulobacter crescentus*, in that they have a common requirement for a novel protein, GmpA. In the absence of GmpA, the chromosomal origins become dissociated from the pole, exhibiting a wobbling motion as if it were no longer fixed in position. Furthermore, stalk formation is inhibited, two histidine kinase signaling proteins fail to localize to the pole, and cell division is defective. GmpA is a cytoplasmic protein that co-localizes with multiple proteins at the cell poles. Biochemical analyses reveal that GmpA self-assembles into filament-shaped polymers that physically interact with several other polar proteins, including the ParB protein at the chromosome origins and two polar histidine kinases. Consistent with this, GmpA accumulates into a large complex at the cell pole when overexpressed, creating a structure that is easily distinguishable from normal cytoplasm by cryo-electron tomography. Those proteins that are mislocalized in the absence of GmpA are recruited to this structure, and the distribution of origin regions is

also altered to reflect the change in GmpA distribution, suggesting the formation of an enlarged polar domain. We conclude that GmpA is responsible for large-scale organization of the *Caulobacter* cell, acting as a key structural component in the assembly of polar multiprotein complexes and the anchoring of the replication origin.

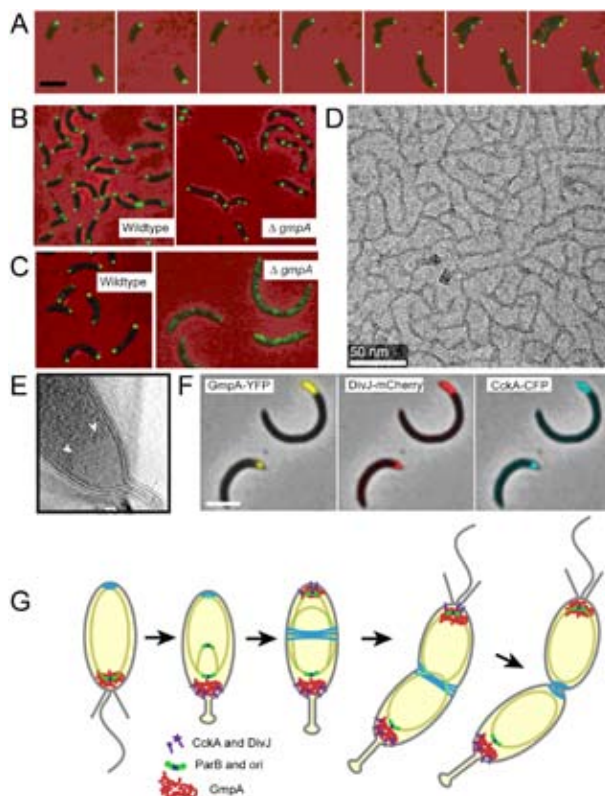


Figure Legend:

(A) GmpA-YFP is localized to the cell poles. Swarmer cells were isolated from a culture of *gmpA-yfp* strain and placed on an agarose pad for microscopic analysis. Overlays of fluorescence (green) and phase contrast images (red shading) at selected times between 0 and 204 minutes are shown.

(B) GmpA anchors the chromosome origin to the cell poles. Here, the origin is marked by CFP-ParB, and its localization is observed in cells expressing GmpA at normal levels (left panel) and $\Delta gmpA$ cells (right panel) CFP-ParB fluorescence (green) is overlaid on the phase contrast image (red shading).

(C) The CckA histidine kinase is mislocalized in the absence of GmpA. CckA-GFP was observed in the presence (left panel) or absence (right panel) of GmpA expression. The fluorescence signal (green) is overlaid on the phase contrast images (red shading).

* Presenting author

(D) GmpA self-assembles into a structured oligomer. GmpA was expressed and purified from *E. coli* cells, placed on a carbon coated grid, and viewed at 86,000X magnification by transmission electron microscopy.

(E) GmpA overexpression induces the formation of large structures at the cell pole. Here, a 1-voxel thick slice from a tomographic reconstruction of a GmpA overexpressing cell is shown, demonstrating a translucent “plug” at a stalked pole. Arrowheads indicate the transition between the plug and normal cytoplasm.

(F) GmpA recruits transmembrane histidine kinase signaling proteins to the cell poles. CckA-GFP and DivJ-mCherry co-localize with the plug formed by the overexpression of GmpA-YFP. Fluorescent images are overlaid on phase contrast (grey shading).

(G) A model of GmpA function. GmpA (in red) accumulates at the new pole at the swarmer to stalked cell transition, and anchors the newly replicated chromosome origin complex (green and blue) to this location. GmpA also serves as an attachment site for histidine kinases (blue and purple) which are part of the control mechanism for chromosome replication in the next round of cell division.

This work was funded by the DOE GTL project titled “Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis”. The PI of this multiinstitutional grant is Harley McAdams, Stanford University.

96 RAPTOR: Robust Alignment and Projection Estimation for TOMographic Reconstruction

—
GTL

Fernando Amat^{1*} (famat@stanford.edu), Farshid Moussavi,¹ Kenneth H. Downing,² Luis R. Comolli,² Mark Horowitz,¹ and **Harley McAdams**³ (hmcadams@stanford.edu)

¹Department of Electrical Engineering, Stanford University, Stanford, California; ²Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; and ³Department of Developmental Biology, Stanford University, Stanford, California

Project Goals: Image enhancement in electron microscopy tomography. Use of statistical imaging and reasoning techniques to improve quality of electron

microscopy images. Perform quantitative analysis of the 3D images to acquire biological insights.

We present a method for automatic full-precision alignment of the images in a tomographic tilt series. Full-precision automatic alignment of cryo electron microscopy images has remained a difficult challenge to date, due to the limited electron dose and low image contrast. These facts lead to poor signal to noise ratio (SNR) in the images, which causes automatic feature trackers to generate errors, even with high contrast gold particles as fiducial features. To enable fully automatic alignment for full-precision reconstructions, we frame the problem probabilistically as finding the most likely particle tracks given a set of noisy images, using contextual information to make the solution more robust to the noise in each image. To solve this maximum likelihood problem, we use Markov Random Fields (MRF) to establish the correspondence of features in alignment. The resulting algorithm, called Robust Alignment and Projection Estimation for Tomographic Reconstruction, or RAPTOR, has not needed any manual intervention for the difficult datasets we have tried, and has provided sub-pixel alignment that is as good as the manual approach by an expert user. Our method has been applied to challenging cryo electron tomographic datasets with low SNR from intact bacterial cells, as well as several plastic section and X-ray datasets.

Currently, the software is freely available on the web and several labs (LBL, UCSD, Caltech) are using it for tomographic reconstructions. This work has recently been published in the Journal of Structural Biology in [1].

This work was funded by the DOE GTL project titled “Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis”. The PI of this multiinstitutional grant is Harley McAdams, Stanford University.

Reference

1. F. Amat and F. Moussavi and L.R. Comolli and G. Elidan and K.H. Downing and M. Horowitz. *Markov random field based automatic image alignment for electron tomography*. Journal of Structural Biology, 2007 (In press).

* Presenting author

97

GTL

Genome-Wide Analysis of Polarly Localized Protein Complexes in *Caulobacter crescentus*: Function, Composition and Spatio-Temporal Dynamics

Edgar Huitema, Sean Pritchard, Sunish Kumar Radhakrishnan, and **Patrick H. Viollier*** (phv1@case.edu)

Department of Molecular Biology and Microbiology, School of Medicine, Case Western Reserve University, Cleveland, Ohio

Project Goals: Identification of polar multiprotein complexes in *Caulobacter crescentus* using a combination of genome-wide genetic and cell biological approaches.

Genetic and cell biological screens were conducted on a genome-wide scale to unearth novel multi-protein complexes that are sequestered to the cell pole in *Caulobacter crescentus*, a prokaryotic model organism for the study of cell polarity. New transposon- and/or plasmid-based tools were developed, furnished with different genetic tags (encoding fluorescence proteins such as mCherry, superfolder GFP or Venus) either for cytological studies or biochemical analysis (affinity purification tags such as TAP), and subsequently used in random genetic screens to isolate (i) strains with translational fusions to polarly localized proteins and (ii) mutants with polar organizational defects. This approach implicated >100 candidates as components of polar multiprotein complexes and/or in controlling the formation of a (sub)complex at the pole. While efforts to define the precise function, composition and spatio-temporal properties of these complexes are still ongoing, two in-depth studies on such protein assemblages localized either at the younger or the older cell pole, were recently completed^(1,2).

Our first study described the role of the TipN scaffolding protein in orchestrating the assembly of the flagellar nanomachine specifically at the younger cell pole⁽¹⁾. TipN essentially functions as a molecular beacon that marks the correct pole as flagellum assembly site and that recruits building blocks of the flagellum to this pole. TipN is itself already localized to this site when the new pole is formed during cytokinesis. Using cytochemistry, we found TipN to be associated with the cytokinetic machinery that assembles at the division plane and we showed that TipN subsequently remains at the new pole (the flagellum assembly site) that emerges with the

completion of cell division, where it directs flagellum construction in the daughter cell. This study provided the first conclusive evidence of an intimate connection between the cell division machinery and the subsequent formation of a multiprotein assemblage at the new cell pole.

Recently, we also reported a multi-protein complex that is specifically associated with the older of the two cell poles⁽²⁾. Briefly, we discovered that a developmental protein kinase, DivJ, which regulates the asymmetric division cycle of *Caulobacter* is recruited to the old cell pole by SpmX, a protein homologous to cell wall hydrolytic enzymes. SpmX is required to stimulate DivJ kinase activity and resides in a complex with DivJ, indicating that this class of enzymes, previously believed to have structural roles in cell wall metabolism, can also have profound regulatory functions. SpmX localizes to the old pole prior to DivJ and the enzymatic domain of SpmX is required and sufficient for polar localization. This suggests that organizational signals might exist in the cell wall to nucleate protein assemblages at the poles.

References

1. Huitema E, Pritchard S, Matteson D, Radhakrishnan SK, and Viollier PH. (2006) Bacterial birth scar proteins mark future flagellum assembly site. *Cell*, 124(5): 1025-37.
2. Radhakrishnan SK, Thanbichler M, and Viollier PH. (2008). The dynamic interplay between a lysozyme homolog and a cell fate determinant drives the asymmetric division cycle of *Caulobacter crescentus*. *Genes & Development*, 22(2): *in press*.

This work was funded by the DOE GTL project titled "Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis". The PI of this multi-institutional grant is Harley McAdams, Stanford University.

* Presenting author

New Methods for Whole-Genome Analysis of Protein Localization

J. Werner, J. Guberman, E. Chen, A. Zippilli, and
Z. Gitai* (zgitai@princeton.edu)

Department of Molecular Biology, Princeton
University, Princeton, New Jersey

Project Goals: To characterize bacterial cell architecture through developing and implementing methods for analyzing the subcellular localization of whole proteomes.

In the past decade, the emerging field of bacterial cell biology has established the importance of properly distributing and organizing subcellular components. It is now clear that virtually every known cellular process requires the precise coordination of the right components to the right cellular address at the right time. To date, however, the labor involved in these studies has required the field to focus on the localization of a small set of previously-characterized proteins. To extend the powerful analysis of protein localization to a whole-genome scale, we have established a series of high-throughput methods for generating, imaging, and analyzing fluorescent protein fusions. Using these new methods, we have determined the localization of over 3,250 proteins in the aquatic bacterium, *Caulobacter crescentus*, and have identified over 450 proteins with distinct non-uniform localizations. We are now in the process of using this data to both deepen our understanding of the functions of specific proteins, as well as to broaden our grasp of the systems-level spatial and temporal organization of a bacterial cell. Our discovery of multiple metabolic pathways with distinct localizations has significant implications for future metabolic engineering efforts. In addition, our high-throughput methods have accelerated our ability to functionally annotate a large number of proteins of previously-unknown function. For example, we have used these methods to identify novel regulators of *Caulobacter* cell division. This poster will discuss the methods we have developed, how they can be adapted to other species of interest, and the systems-level and gene-specific discoveries we have already made.

This work was funded by the DOE GTL project titled “Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis”. The PI of this multiinstitutional grant is Harley McAdams, Stanford University.

* Presenting author