

# Systems Environmental Microbiology

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

### ESPP Functional Genomics and Imaging Core (FGIC): Cell Wide Analysis of Metal-Reducing Bacteria

Aindrila Mukhopadhyay,<sup>1,6\*</sup> Edward Baidoo,<sup>1,6</sup> Kelly Bender<sup>5,6</sup> (bender@micro.siu.edu), Peter Benke,<sup>1,6</sup> Swapnil Chhabra,<sup>1,6</sup> Elliot Drury,<sup>3,6</sup> Masood Hadi,<sup>2,6</sup> Zhili He,<sup>4,6</sup> Jay Keasling<sup>1,6</sup> (keasling@berkeley.edu), Kimberly Keller,<sup>3,6</sup> Eric Luning,<sup>1,6</sup> Francesco Pingitore,<sup>1,6</sup> Alyssa Redding,<sup>1,6</sup> Jarrod Robertson,<sup>3,6</sup> Rajat Sapra,<sup>2,6</sup> Anup Singh<sup>2,6</sup> (aksingh@sandia.gov), Judy Wall<sup>3,6</sup> (wallj@missouri.edu), Grant Zane,<sup>3,6</sup> Aifen Zhou,<sup>4,6</sup> and Jizhong Zhou<sup>4,6</sup> (jzhou@rccc.ou.edu)

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>2</sup>Sandia National Laboratories, Livermore, Calif.; <sup>3</sup>University of Missouri, Columbia, Mo.; <sup>4</sup>University of Oklahoma, Norman, Okla.; <sup>5</sup>Southern Illinois University, Carbondale, Ill.; and <sup>6</sup>Virtual Institute of Stress and Survival (vimss.lbl.gov).

**Project Goals: The primary goal of the Environmental Stress Pathway Project (ESPP) is a rigorous understanding of the sulfate reducing bacterium (SRB), *Desulfovibrio vulgaris* Hildenborough physiology and its ability to survive in its environment.**

The primary goal of the Environmental Stress Pathway Project (ESPP) is a rigorous understanding of the sulfate reducing bacterium (SRB), *Desulfovibrio vulgaris* Hildenborough physiology and its ability to survive in its environment. This knowledge provides the basis for discerning the biogeochemistry at metal contaminated sites, for bioremediation and natural attenuation for toxic metals. The FGIC focuses on mapping these responses at a cell wide level using systems biology approaches. In the last one year, our methods that have been optimized and utilized over the years to study a variety of growth/ stress conditions and mutants, were extended to study more environmentally relevant physiological conditions and the interaction of *D. vulgaris* with other microbes.

Generating high quality biomass continues to be a critical aspect for all our functional genomics studies. This was found to be especially important for complex conditions such as biofilm formation, long term exposure to stress and the study of mixed cultures containing multiple organisms. These studies were conducted in close collaboration with the Applied and Environmental Core and additional methods development had to be undertaken to address

GTL

many unique aspects of studying such complex systems. For example, the study of the *D. vulgaris* / *Methanococcus marisnigri* syntrophic co-culture (serving as a model of naturally occurring SRB/ Methanogen interactions) required optimization of microarray hybridization methods and an alternate workflow for iTRAQ proteomics application. Our team also has advanced tools for metabolite level analysis, such as a <sup>13</sup>C isotopomer based flux analysis which provides valuable information about bacterial physiology. However the study of individual organisms in a mixed culture using existing flux analysis methods is difficult since the method typically relies on amino acids from hydrolyzed proteins from a homogenous biomass. To overcome the need to separate the target organism in a mixed culture, we successfully explored the idea that a single highly-expressed protein could be used to analyze the isotopomer distribution of amino acids from one organism. An overview of these studies and key observations are presented.

Additionally we continued to collect cell wide data in *Shewanella oneidensis* and *Geobacter metallireducens* for comparative studies. Improved methods for extraction and high throughput of metabolite analysis using CE-MS and LC-MS were applied to several studies underway in ESPP, many of which are also required quantitation. In collaboration with Computational Core, a novel FTICR-MS based method for a comparative <sup>12</sup>C/<sup>13</sup>C based metabolite analysis was also developed to enable a direct comparison of control and experimental samples for relative quantitation.

Improved methods for generating stable knockout mutants and marker-less mutants in *D. vulgaris* has been now been widely used to follow up hypothesis from a majority of our stress response and most valuably in confirming candidates for “missing steps” in *D. vulgaris* metabolism. Sets of targeted mutants are also being constructed to study the large number of two component signaling systems in *D. vulgaris*. To ensure a complete understanding of regulatory mechanisms, the study alternative regulatory mechanisms such as small non-coding RNAs are also underway. In collaboration with the computational core, work is in progress to set up searchable databases of all our large data sets, including proteomics, metabolite and flux data.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Analysis of a *Desulfovibrio vulgaris* Small RNA and Its Target Under Various Stress Conditions

Andrew S. Burns<sup>1,2</sup> and Kelly S. Bender<sup>1,2\*</sup> (bender@micro.siu.edu)

<sup>1</sup>Microbiology Dept., Southern Illinois University, Carbondale, Ill.; and <sup>2</sup>VIMSS (Virtual Institute for Microbial Stress and Survival), <http://vimss.lbl.gov>

**Project Goals: Our goal is to understand the intricate regulatory cascades involved in how the model metal reducer *Desulfovibrio vulgaris* and its relatives respond to stressors in contaminated sites.**

One of the aims of the Environmental Stress Pathway Project is to elucidate regulatory networks critical to processes of interest to the DOE. As such, our goal is to understand the intricate regulatory cascades involved in how the model metal reducer *Desulfovibrio vulgaris* and its relatives respond to stressors in contaminated sites. One approach we are taking to meet this challenge is the identification and analysis of small non-coding RNA molecules (sRNAs). Ranging in size from 20-200 nucleotides, 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. Here we report the analysis of Dv-sRNA2, a molecule previously identified from a random small RNA clone library, and its target gene DVU0678- a hypothetical protein only present in the *D. vulgaris* strains Hildenborough and DP4.

While expression of Dv-sRNA2 has been confirmed, its regulatory role is currently unknown. Dv-sRNA2 is located in the same chromosomal region as its putative target DVU0678, but on the opposite strand. As such, a Dv-sRNA2 deletion mutant cannot be constructed without affecting the expression of DVU0678. To circumvent this problem, the gene encoding Dv-sRNA2 was cloned into the stable vector pMO719 to elucidate the effects of over-expression. This construct resulted in a strain (KB100) containing two copies of Dv-sRNA2 under the control of their native promoter. Phenotypic analysis of the KB100 strain compared to a control strain harboring an empty vector indicated no difference in growth under normal 37°C/pH 7 and pH 6 growth conditions. However, a slight increase in growth rate was observed for KB100 when grown at 45°C as well as when 45°C growth was shifted to 50°C during early log-phase. This minimal phenotypic difference likely resulted from similar expression rates of Dv-sRNA2 in strain KB100 and the control. Since it appeared that the merodiploid strain was not over-expressing the sRNA gene, a strain was constructed in which the entire Dv-sRNA2/DVU0678 region was deleted via maker exchange. Analysis of this deletion strain (KB102) compared to the wild type indicated similar growth patterns during 25°C and 37°C growth at pH 7. However, increasing the growth temperature to 45°C increased the lag phase of KB102 compared to the wild type by nine hours. While

these data implicate that Dv-sRNA2 and DVU0678 are involved in the *D. vulgaris* heat shock response, studies involving other stressors are currently underway. New strategies for constructing over-expression strains are also being developed to help predict the role individual sRNAs have on the physiology and transcriptional response of *D. vulgaris* under multiple environmental conditions.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Functional Characterization of Microbial Genomes by Tagged Transposon Mutagenesis

Adam Deutschbauer<sup>1,4\*</sup> (AMDeutschbauer@lbl.gov), Julia Oh,<sup>2</sup> Morgan Price,<sup>1,4</sup> Jennifer Kuehl,<sup>1,4</sup> Paramvir Dehal,<sup>1,4</sup> Corey Nislow,<sup>3</sup> Ronald W. Davis,<sup>2</sup> and Adam P. Arkin<sup>1,4</sup>

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

**Project Goals: A primary goal of the Environmental Stress Pathway Project (ESPP) is a systems-level model of sulfate-reducing bacteria (SRB) metabolism, stress responses, and gene regulation. However, current systems-level analyses of less studied bacteria such as SRBs are limited by the presence of numerous uncharacterized genes and an over reliance on annotations from well studied bacteria such as *E. coli*.**

A primary goal of the Environmental Stress Pathway Project (ESPP) is a systems-level model of sulfate-reducing bacteria (SRB) metabolism, stress responses, and gene regulation. However, current systems-level analyses of less studied bacteria such as SRBs are limited by the presence of numerous uncharacterized genes and an over reliance on annotations from well studied bacteria such as *E. coli*. Therefore, it is imperative that rapid and quantitative methods are developed to determine microbial gene function in a high-throughput manner. 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. We have cloned and sequence-verified 4280 tag modules into a Gateway entry vector. Each tag 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 tag 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 4280 tag modules was used and which gene was disrupted. Transposon mutants can be rapidly re-arrayed into a single pool containing 4280 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 by the hybridization of the barcodes to an Affymetrix microarray containing the tag complements in a system identical to that used for the yeast deletion collection. To facilitate both strain construction and mutant pool phenotyping, we have implemented a robotic infrastructure (both aerobic and anaerobic) for assay setup and automated mutant pool growth, collection, and processing. Here we describe the initial application of our approach to the metal-reducing bacterium *Shewanella oneidensis* MR1 and the SRBs *Desulfovibrio desulfuricans* G20 and *Desulfovibrio vulgaris* Hildenborough.

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, provide insight into the functional connectivity of different pathways, and enable the construction of genome-wide models of SRB function and activity.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

—  
GTL

## The Development and Application of an Integrated Functional Genomics Platform in *Desulfovibrio desulfuricans* G20

Adam Deutschbauer,<sup>1,2</sup> Jennifer Kuehl<sup>1,2\*</sup> (JKuehl@lbl.gov), Morgan Price,<sup>1,2</sup> Paramvir Dehal,<sup>1,2</sup> Terry C. Hazen,<sup>1,2</sup> and Adam P. Arkin<sup>1,2</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>2</sup>VIMSS (Virtual Institute for Microbial Stress and Survival), <http://vimss.lbl.gov>

**Project Goals: We are developing an integrated functional genomics platform in the SRB *Desulfovibrio desulfuricans* G20 (G20). The platform is comprised of phenotypic arrays, gene expression arrays, and a transposon mutant library. A hallmark of our transposon library is the ability to pool thousands of mutants for parallel phenotypic analysis. Profiling our transposon library under a large number of conditions will enable us to link a large fraction of the G20 genome to specific cellular phenotypes. Here**

## we present our experimental approaches and preliminary data from each type of experiment.

The Environmental Stress Pathway Project (ESPP) focuses on the systems-biology and environmental activity of the sulfate-reducing bacterium (SRB) *Desulfovibrio vulgaris* Hildenborough (DvH). However, focusing on a single representative species of a genus offers little insight into the evolution, niche adaptation, and function of an entire microbial genus. Given the broad environmental significance of SRBs and their phylogenetic diversity, it is imperative that other representative species of the *Desulfovibrio* genus are investigated. Therefore, we are developing an integrated functional genomics platform in the SRB *Desulfovibrio desulfuricans* G20 (G20). The platform is comprised of phenotypic arrays, gene expression arrays, and a transposon mutant library. The ability of G20 to utilize carbon sources and electron acceptors for growth, survive stress conditions, and respond to growth inhibitors is measured on a high-throughput system to narrow the test conditions to be used in both gene expression and mutant profiling assays. Using high-density multiplex microarrays, we are monitoring gene expression under a number of diverse conditions in a global effort to elucidate G20 gene regulation. Our principle effort is the construction of a genome-wide, tagged transposon mutant library in G20. A hallmark of our transposon library is the ability to pool thousands of mutants for parallel phenotypic analysis. Profiling our transposon library under a large number of conditions will enable us to link a large fraction of the G20 genome to specific cellular phenotypes. Here we present our experimental approaches and preliminary data from each type of experiment.

Functional genomic investigations into G20 will be used by the ESPP in a number of ways. First, G20 and DvH have a large number of orthologous genes. Therefore, functional insights made in G20 are a strong starting point for targeted experiments in DvH. Second, by expanding our experimental “pan-genome” to include G20, we are now investigating more of the natural variation present in the genus *Desulfovibrio*. These data will be imperative for annotating and analyzing newly sequenced SRB genomes. Lastly, our integrated functional genomics data will be used to build a cellular model of G20 metabolism, regulation, and activity. This model can be used as a framework for the interpretation of many types of ESPP data generated in the laboratory and in the field.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

## A Phylogenomic Approach to the Evolutionary Origins of Microbial Metabolisms

Lawrence David<sup>1\*</sup> (ldavid@mit.edu) and Eric Alm<sup>1,2</sup>

<sup>1</sup>Massachusetts Institute of Technology, Cambridge, Mass.; and <sup>2</sup>The Virtual Institute for Microbial Stress and Survival, Berkeley, Calif.

**Project Goals: This work is consistent with the DOE stated goal “to gain insights about fundamental biological processes.”**

We have developed a tool, AnGST, to infer the evolutionary history of gene families independent of species phylogenies. This allows for ‘birth dates’ to be assigned to individual genes. We use this genetically encoded history of functional genetic material to infer the types of molecular functions present on the Earth as a function of time. For example, enzymes utilizing molecular oxygen are found to be largely absent until ~2.7 Gya, consistent with geological estimates of the origins of oxygenic photosynthesis.

GTL

## Resource for the Exploration of Regulons Accurately Predicted by the Methods of Comparative Genomics

Pavel S. Novichkov<sup>1\*</sup> (PSNovichkov@lbl.gov), Mikhail S. Gelfand,<sup>2</sup> Inna Dubchak,<sup>1,4</sup> and Dmitry A. Rodionov<sup>2,3</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>2</sup>Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia; <sup>3</sup>Burnham Institute for Medical Research, La Jolla, Calif.; and <sup>4</sup>DOE Joint Genome Institute, Walnut Creek, Calif.

**Project Goals: The current version of database covers more than 250 genomes and 180 profiles. Among others, it represents the results of our recent comparative genomic reconstruction of metabolic regulons in 13 *Shewanella* species that included near 70 transcription factors, approximately 400 binding sites and more than 1000 target genes per each genome. The database gives access to large regulatory networks reconstructed for certain metabolic pathways, e.g. degradation of fatty acids, branch chain amino acids, and aminosugars, homeostasis of biometals, and biosynthesis of NAD cofactor. In the near future we are planning to add a large collection of regulons for the LacI family transcription factors.**

Identification and reconstruction of various transcriptional regulons in bacteria using a computational comparative genomics approach is coming of age. During the past decade a large number of manually-curated high quality inferences of transcriptional regulatory interactions were accumulated

for diverse taxonomic groups of bacteria. These data provide a good foundation for understanding molecular mechanisms of transcriptional regulation, identification of regulatory circuits, and interconnections among circuits within the cell. Traditional experimental methods for regulon analysis have certain limitations both in terms of productivity and feasibility. While the development of high-throughput transcriptome approaches allow to obtain genome-scale gene expression patterns, in many cases the complexity of the interactions between regulons makes it difficult to distinguish between direct and indirect effects on transcription. The availability of a large number of closely related genomes allows one to apply comparative genomics to accurately expand already known regulons to yet uncharacterized organisms, and to predict and describe new regulons. Due to fast accumulation of such valuable data, there is a need for a specialized database and associated analysis tools that will compile and present the growing collection of high quality predicted bacterial regulons.

The RegPrecise database was developed for capturing, visualization and analysis of transcription factor regulons that were reconstructed by the comparative genomic approach. The primary object of the database is a single regulon in a particular genome, which is described by the identified transcription factor, its DNA binding site model (a profile), as well as the set of regulated genes, operons and associated operator sites. Regulons for orthologous transcription factors from closely related genomes are combined into the collections that provide an overview of the conserved and variable components of the regulon. A higher level representation of the regulatory interactions is also provided for orthologous regulons described in several bacterial taxonomic groups enabling comparison and evolutionary analysis of the transcription factor binding motifs. Another view of complex data in the database is a general overview of multiple regulons inferred in a set of closely related group of genomes.

The current version of database covers more than 250 genomes and 180 profiles. Among others, it represents the results of our recent comparative genomic reconstruction of metabolic regulons in 13 *Shewanella* species that included near 70 transcription factors, approximately 400 binding sites and more than 1000 target genes per each genome. The database gives access to large regulatory networks reconstructed for certain metabolic pathways, e.g. degradation of fatty acids, branch chain amino acids, and aminosugars, homeostasis of biometals, and biosynthesis of NAD cofactor. In the near future we are planning to add a large collection of regulons for the LacI family transcription factors.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Expression Profiling of Hypothetical Genes in *Desulfovibrio vulgaris* Leads to Improved Functional Annotation

Dwayne A. Elias\* (eliasd@missouri.edu) Elliott C. Drury, Alyssa M. Redding, Aindrila Mukhopadhyay, Marcin Joachimiak, Huei-Che B. Yen, Matthew W. Fields,<sup>3</sup> Terry C. Hazen,<sup>1</sup> **Adam P. Arkin** (APArkin@lbl.gov), Jay D. Keasling,<sup>1</sup> and Judy D. Wall<sup>2</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.;

<sup>2</sup>University of Missouri, Columbia, Mo.; and <sup>3</sup>Montana State University, Bozeman, Mont.

**Project Goals: Hypothetical and conserved hypothetical genes account for >30% of sequenced bacterial genomes. For the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough, 348 of the 3534 genes are annotated as conserved hypothetical (9.7%) with 889 hypothetical genes (25.0%). Given this large genome fraction, it is plausible that some genes serve critical cellular functions. The goals of this study were to determine which genes can be expressed and to provide a more functionally based annotation.**

Hypothetical and conserved hypothetical genes account for >30% of sequenced bacterial genomes. For the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough, 348 of the 3534 genes are annotated as conserved hypothetical (9.7%) with 889 hypothetical genes (25.0%). Given this large genome fraction, it is plausible that some genes serve critical cellular functions. The goals of this study were to determine which genes can be expressed and to provide a more functionally based annotation. To accomplish this, expression profiles of the 1237 hypothetical and conserved hypothetical genes were obtained from transcriptomic datasets of 10 environmental stresses, complimented with iTRAQ proteomic data. Genes were divided into putatively polycistronic operons and those predicted to be monocistronic, then classified by basal expression levels and grouped according to changes in expression for one or multiple stresses. 1219 of these genes were transcribed with 265 proteins detected. There was no evidence for expression of 17 predicted genes. Except for the latter, annotation of all monocistronic genes was expanded using the above criteria and COG information. Polycistronic genes were annotated with expression information including proximity to more confidently annotated genes. Two targeted deletion mutants were used as test cases to determine the accuracy of the inferred functional annotations.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Impact of Elevated Nitrate on Sulfate-Reducing Bacteria: Implications of Inhibitory Mechanisms in Addition to Osmotic Stress

Qiang He,<sup>1,2\*</sup> Zhili He,<sup>3,13</sup> Dominique C. Joyner,<sup>4,13</sup> Marcin Joachimiak,<sup>5,13</sup> Morgan N. Price,<sup>5,13</sup> Zamin K. Yang,<sup>6,13</sup> Huei-Che Bill Yen,<sup>7,13</sup> Christopher L. Hemme,<sup>3,13</sup> Romy Chakraborty,<sup>4,13</sup> Wenqiong Chen,<sup>8,13</sup> Matthew M. Fields,<sup>9,13</sup> David A. Stahl,<sup>12,13</sup> Jay D. Keasling,<sup>5,11,13</sup> Martin Keller,<sup>6,13</sup> **Adam P. Arkin**<sup>5,10,13</sup> (APArkin@lbl.gov), Terry C. Hazen,<sup>4,13</sup> Judy D. Wall,<sup>7,13</sup> and Jizhong Zhou<sup>3,13</sup>

<sup>1</sup>Dept. of Civil and Environmental Engineering and <sup>2</sup>Center for Environmental Biotechnology, University of Tennessee, Knoxville, Tenn.; <sup>3</sup>Institute for Environmental Genomics, Dept. of Botany and Microbiology, University of Oklahoma, Norman, Okla.; <sup>4</sup>Earth Sciences Division and <sup>5</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>6</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>7</sup>Depts. of Biochemistry and Molecular Microbiology & Immunology, University of Missouri, Columbia, Mo.; <sup>8</sup>Diversa Corp, San Diego, Calif.; <sup>9</sup>Dept. of Microbiology, Montana State University, Bozeman, Mont.; <sup>10</sup>Dept. of Bioengineering and <sup>11</sup>Dept. of Chemical Engineering, University of California, Berkeley, Calif.; <sup>12</sup>Dept. of Civil and Environmental Engineering, University of Washington, Seattle, Wash.; and <sup>13</sup>Virtual Institute for Microbial Stress and Survival

Sulfate-reducing bacteria are studied for their potential in heavy metal bioremediation. However, the occurrence of elevated nitrate in contaminated environments has been shown to inhibit sulfate reduction activity. While the inhibition has been suggested to result from competition with nitrate-reducing bacteria, the possibility of direct inhibition of sulfate reducers by elevated nitrate needs to be explored. Using *Desulfovibrio vulgaris* as a model sulfate-reducing bacterium, it was observed that significant growth inhibition was effected by 70 mM NaNO<sub>3</sub> but not 70 mM NaCl, indicating the presence of inhibitory mechanisms in addition to osmotic stress. While the differential expression of a small number of genes in response to nitrate suggested the potential involvement of osmotic and nitrite stress responses, the roles of these two stress responses appear minor given the lack of similarity in the overall transcriptional profiles between nitrate, nitrite, and NaCl stress responses. The presence of unique stress response pathways in nitrate stress is further suggested by the lack of extensive similarities in the response profiles between nitrate stress and various other stress conditions. In addition, the importance of genes with functions in the metabolism of S-adenosylmethionine in the shift of energy flow was implicated in nitrate stress response.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through

contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

## A Role of CO and a CO Sensor Protein in the Energy Metabolism of *D. vulgaris* Hildenborough

Lara Rajeev<sup>1\*</sup> (lrajeev@u.washington.edu), Sergey Stolyar,<sup>1</sup> David A Stahl,<sup>1</sup> Grant Zane,<sup>2</sup> and Judy Wall<sup>2</sup>

<sup>1</sup>University of Washington, Seattle, Wash.; and <sup>2</sup>University of Missouri, Columbia, Mo.

**Project Goals: We put forth the hypothesis that Coo and Hyn1 (NiFe) hydrogenases are involved in hydrogen production from lactate and pyruvate, and in the absence of either hydrogenase, electrons from lactate and pyruvate are directed towards CO production.**

Recent studies suggest that carbon monoxide (CO) may play an important role in the energy metabolism of sulfate-reducing bacteria. The membrane bound cytoplasmically oriented CooMKLUXHF hydrogenase of *D. vulgaris* Hildenborough is similar to the Coo hydrogenases of *R. rubrum* and *C. hydrogenoformans*. In the latter systems, the Coo hydrogenases together with the CO dehydrogenase (CODH) oxidize CO to CO<sub>2</sub> and H<sub>2</sub>. Both operons are regulated by a CO sensing transcriptional regulator CooA. The *D. vulgaris* Hildenborough genome also encodes genes for CODH and CooA. Predicted binding sites for CooA are located upstream of both CODH and the Coo hydrogenase operons. To determine if DvH CooA also acts as a CO sensor and if Coo hydrogenase is CO-regulated, we tested a mutant deleted for the *cooA* gene for growth on lactate-sulfate (LS) or pyruvate-sulfate (PS) in the presence of CO. With 1% CO in the headspace, wild type DvH grows efficiently and consumes CO, whereas a *cooA* mutant does not oxidize the CO. Interestingly, the *cooA* mutant grew efficiently on PS with CO in the headspace, but on LS the growth was poor and was inhibited by high H<sub>2</sub> accumulation. The *cooL* mutant lacking an active Coo hydrogenase was able to consume CO. Both wt DvH and the *cooL* mutant, but not the *cooA* mutant, were able to grow on CO as the sole energy source, although growth was very slow. The *cooA* mutant could grow syntrophically on lactate with *M. maripaludis*, whereas the *cooL* mutant could not grow. These observations suggest that the *coo* hydrogenase expression is not CO- or CooA-dependent, unlike that seen in the *R. rubrum* system.

To determine the physiological role of CO in DvH metabolism, we followed the fermentation burst in various DvH hydrogenase mutants. Wt DvH produces very little CO during growth. However, the *cooL* mutant shows a pronounced CO burst during growth on both LS and PS. Other mutants lacking either the Ech hydrogenase or the Hyd (Fe) hydrogenase showed no CO burst, but the *hyd hyn1* double mutant showed a CO burst. Notably, the *cooA* mutant

did not accumulate CO during growth. We put forth the hypothesis that Coo and Hyn1 (NiFe) hydrogenases are involved in hydrogen production from lactate and pyruvate, and in the absence of either hydrogenase, electrons from lactate and pyruvate are directed towards CO production.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

## Comparison of the Sulfate-Reducing Capacity of *Desulfovibrio vulgaris* Hildenborough Deleted for the Operon Containing *qmoABC* and a Hypothetical Protein (DVU0851) versus Deletion of the Hypothetical Protein Alone

Grant M. Zane,\* Dwayne A. Elias, Huei-Che Yen, and Judy D. Wall

Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>)

**Project Goals: Comparison of the sulfate-reducing capacity of *Desulfovibrio vulgaris* Hildenborough deleted for the operon containing *qmoABC* and a hypothetical protein (DVU0851) versus deletion of the hypothetical protein alone.**

Deletion of the operon encoding *qmoABC* and a hypothetical protein (HP) (DVU0848-51) in the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough resulted in an inability to respire sulfate. No suppressed mutants appeared in cultures of the deletion strain incubated in the presence of sulfate. Curiously, the  $\Delta(qmoABC\ HP)$  mutant was also unable to ferment pyruvate. Respiration of sulfate and fermentation of pyruvate was restored by complementation with the *qmoABC*, HP genes. In order to determine the contribution of the promoter-distal hypothetical protein to the ability of this organism to reduce sulfate and ferment pyruvate, a second deletion was made of this gene alone. Although the mutant deleted for the single gene was able to reduce sulfate, it grew more slowly than wild-type and was stimulated by methionine and cysteine. Complementation restored growth to near wild-type levels.

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

## The Molecular Mechanism of Adaptation to Salt Stress Revealed by the Long-Term Evolution of *Desulfovibrio vulgaris* Hildenborough

Aifen Zhou,<sup>1,6\*</sup> Zhili He,<sup>1,6</sup> Marcin P. Joachimiak,<sup>2,6</sup> Paramvir S. Dehal,<sup>2,6</sup> Adam P. Arkin<sup>2,6</sup> (APArkin@lbl.gov), Kristina Hillesland,<sup>4,6</sup> David Stahl,<sup>4,6</sup> Judy Wall,<sup>5,6</sup> Terry C. Hazen,<sup>3,6</sup> and Jizhong Zhou<sup>1,6</sup>

<sup>1</sup>Institute for Environmental Genomics, Dept. of Botany and Microbiology, University of Oklahoma, Norman, Okla.; <sup>2</sup>Physical Biosciences Division and <sup>3</sup>Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>4</sup>Civil & Environmental Engineering, University of Washington, Seattle, Wash.; <sup>5</sup>Depts. of Biochemistry and Molecular Microbiology & Immunology, University of Missouri, Columbia, Mo.; and <sup>6</sup>Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov>

**Project Goals: In order to investigate the long-term evolutionary responses, diversifications and adaptation to salt stress, sulfate reducing bacteria *Desulfovibrio vulgaris* Hildenborough, a model environmental organism, is used to address such fundamental questions by mimicking the stress condition in the lab culture.**

One of the greatest challenges in biology is to understand the interaction between genotype and environment to determine the fitness of an organism. With the recent advances in genome sequencing and high-throughput genomic technologies, now it is possible to link sub-cellular molecular/metabolic processes with the population-level processes, functions and evolution. In order to investigate the long-term evolutionary responses, diversifications and adaptation to salt stress, sulfate reducing bacteria *Desulfovibrio vulgaris* Hildenborough, a model environmental organism, is used to address such fundamental questions by mimicking the stress condition in the lab culture. Control lines and stressed lines (6 lines each, from single colony based pure culture) grown in medium LS4D and LS4D + 100 mM NaCl were transferred every 48 hrs with one to one hundred dilutions. Phenotype of all the cell lines in terms of salt tolerance was tested with LS4D supplemented with 250 mM NaCl. Results demonstrated that the adaptation to salt stress is a dynamical process. The enhanced salt tolerance of stressed lines was observed at 300 generations and became more obvious with the increase of generations. The de-adaptation experiment on 500, 1000 and 1200 generation cell lines not only provided strong evidence that the phenotype was due to the genetic change instead of physiological adaptation, but also indicated that there is also a dynamic trend for genetic adaptation and the genetic mutation might become stable at 1000 generation. In order to further understand the molecular mechanisms of adaptation to salt stress in long-term evolution process, gene expression profiles of the 500 and 1000 generation samples were examined by *D. vulgaris*

whole genome oligo microarray. "Energy production and conversion" and "signal transduction mechanisms" are among the gene categories with most genes up-regulated. Statistical analysis also showed that gene expression profiling between evolved lines with salt stress vs control evolved lines and the evolved lines with salt stress vs ancestor are more closer at 1000 generation. Whole genome sequencing on selected colonies is underway to identify the beneficial genetic mutation.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

## *Desulfovibrio vulgaris* Hildenborough Responses to Salt and H<sub>2</sub>O<sub>2</sub> Stresses

Zhili He,<sup>1,10\*</sup> Aifen Zhou,<sup>1,10</sup> Qiang He,<sup>2,10</sup> Aindrila Mukhopadhyay,<sup>3,10</sup> Edward Baidoo,<sup>3,10</sup> Marcin Joachimiak,<sup>3,10</sup> Christopher L. Hemme,<sup>1,10</sup> Peter Benke,<sup>3,10</sup> Alyssa M. Redding,<sup>3,10</sup> Matthew M. Fields,<sup>4,11</sup> David A. Stahl,<sup>5,10</sup> Jay D. Keasling,<sup>3,6,10</sup> Adam P. Arkin<sup>3,7,10</sup> (APArkin@lbl.gov), Terry C. Hazen,<sup>8,10</sup> Judy D. Wall,<sup>9,10</sup> and Jizhong Zhou<sup>1,10</sup>

<sup>1</sup>Institute for Environmental Genomics, Dept. of Botany and Microbiology, University of Oklahoma, Norman, Okla.; <sup>2</sup>Dept. of Civil and Environmental Engineering, University of Tennessee, Knoxville, Tenn.; <sup>3</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>4</sup>Dept. of Microbiology, Montana State University, Bozeman, Mont.; <sup>5</sup>Dept. of Civil and Environmental Engineering, University of Washington, Seattle, Wash.; <sup>6</sup>Dept. of Chemical Engineering, University of California, Berkeley, Calif.; <sup>7</sup>Dept. of Bioengineering, University of California, Berkeley, Calif.; <sup>8</sup>Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>9</sup>Depts. of Biochemistry and Molecular Microbiology & Immunology, University of Missouri, Columbia, Mo.; and <sup>10</sup>Virtual Institute for Microbial Stress and Survival

**Project Goals: Diverse stress resistance mechanisms may be used in *D. vulgaris* for detoxification of H<sub>2</sub>O<sub>2</sub> with the up-regulation of DNA repair systems and the down-regulation of energy metabolism and protein synthesis.**

The response of *Desulfovibrio vulgaris* Hildenborough to salt and H<sub>2</sub>O<sub>2</sub> stresses were examined by physiological, global transcriptional, metabolite, and mutagenesis analyses. The growth of *D. vulgaris* was inhibited by 250 mM NaCl or 1 mM H<sub>2</sub>O<sub>2</sub>. Salt adaptation (long-term NaCl exposure) increased the expression of genes involved in amino acid biosynthesis and transport, electron transfer, hydrogen

oxidation, and general stress responses (e.g., heat shock proteins, phage shock proteins, and oxidative stress response proteins). Genes involved in carbon metabolism, cell motility, and phage structures were decreased in expression. Comparison of transcriptomic profiles of *D. vulgaris* responses to salt adaptation with those of salt shock (short-term NaCl exposure) showed some similarity as well as a significant difference. Metabolite assays showed that glutamate and alanine accumulated under salt adaptation, suggesting that they may be used as osmoprotectants in *D. vulgaris*. Addition of amino acids (glutamate, alanine, tryptophan) or yeast extract to the growth medium relieved salt-related growth inhibition. A conceptual model is proposed to link the observed results to currently available knowledge for further understanding the mechanisms of *D. vulgaris* adaptation to elevated NaCl. Under H<sub>2</sub>O<sub>2</sub> conditions, PerR regulon genes were significantly up-regulated, indicating the importance role of PerR in oxidative stress response. In addition, some Fur regulon genes were also strongly induced. Increased gene expression of thiol-peroxidase genes ahpC as well as thioredoxin reductase and thioredoxin genes indicated the involvement of thiol switch in the oxidative stress response. rbr2 was the only significantly up-regulated H<sub>2</sub>O<sub>2</sub> scavenging enzymes. The oxidative stress response of mutants ΔperR and Δfur demonstrated that ahpC and rbr2 were regulated by both Fur and PerR. The links between the up-regulated genes involved in H<sub>2</sub>O<sub>2</sub> scavenging, protein fate, DNA metabolism and lipid metabolism and the down-regulated genes involved in sulfate reduction, energy production and translation were demonstrated by the gene co-expression network. The proteomics data provided further evidence at the translational level and complemented the transcriptomics data. Taken together, diverse stress resistance mechanisms may be used in *D. vulgaris* for detoxification of H<sub>2</sub>O<sub>2</sub> with the up-regulation of DNA repair systems and the down-regulation of energy metabolism and protein synthesis.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Analysis of an Intact Dissimilatory Sulfite Reductase Protein Complex from *Desulfovibrio vulgaris* Using an Ion Mobility QTOF Analyzer

Ming Dong,<sup>1</sup> Michael Daly,<sup>2</sup> Haichuan Liu,<sup>3</sup> Simon Allen,<sup>3</sup> Evelin Szakal,<sup>3</sup> Steven C. Hall,<sup>3</sup> Susan J. Fisher,<sup>1,3</sup> Terry C. Hazen,<sup>1</sup> Jil T. Geller,<sup>1</sup> Mary E. Singer,<sup>1</sup> Lee L. Yang,<sup>1</sup> Jian Jin,<sup>1</sup> H. Ewa Witkowska,<sup>2\*</sup> and Mark D. Biggin<sup>1</sup> (mdbiggin@lbl.gov)

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>2</sup>Waters Corporation; and <sup>3</sup>University of California, San Francisco, Calif.

**Project Goals:** MS analysis of intact and denatured Dsr complex was performed utilizing HD Synapt mass spectrometer. Three forms of the complex were isolated from a 400 L DvH culture using a 3-step protein fractionation protocol. MS analysis demonstrated that three forms of Dsr complexes differed in stoichiometry: the presence of a hexamer [DsrA]<sub>2</sub>[DsrB]<sub>2</sub>[DsrC]<sub>2</sub>, a pentamer [DsrA]<sub>2</sub>[DsvB]<sub>2</sub>[DsvC] and a tetramer [DsrA]<sub>2</sub>[DsvB]<sub>2</sub>, all measured within an error of less than 0.1%, was demonstrated. No differences in molecular masses of subunits participating in three distinct forms of Dsr complex were detected. However, all three subunits demonstrated discrepancies between the experimental and theoretical molecular masses. Analysis of DsrC subunit revealed the presence of post-translational modifications.

Protein extract was isolated from a 400 L *D. vulgaris* culture grown under optimal conditions. Dsr complexes were purified utilizing three sequential steps of fractionation: ammonium sulfate precipitation, anion exchange chromatography (Mono Q) and size exclusion chromatography (Superdex 200). The purity of the preparation was 95%, based on SDS PAGE. Intact complexes in 30 mM ammonium bicarbonate were infused by nanospray into a Synapt HD MS instrument that combines ion mobility, quadrupole and TOF analyzers. In addition, Dsr complexes were analyzed under denaturing conditions utilizing on-line LC MS and MS/MS on the same instrument.

Dsr was detected in three distinct Mono Q column fractions: two of the early-eluting fractions contained all three subunits (DsrA, DsrB and DsrC) while DsrC polypeptide was absent in the late-eluting fraction. No differences in molecular masses of polypeptide constituents participating in the three different complex forms were revealed by ESI MS, within method error approximating 10 ppm. However, all three polypeptides, DsrA, DsrB and DsrC, demonstrated a significant discrepancy between the theoretical and experimental molecular masses suggesting the presence of post translational modifications (PTMs) and/or sequence errors. To explain this discrepancy, DsrC was analyzed in detail. Gas phase fragmentations in the source produced top-down sequence information that demonstrated that the



regions [1-36] and [84-104] were consistent with the DNA sequence. In addition, the presence of a disulfide bridge between Cys 92 and Cys 103 was revealed. Thus the above results pointed to the region [37-83] as a site of putative PTMs. The presence of either trimethylation or acetylation on two Lys residues was demonstrated by LC MS/MS analysis of tryptic peptides derived from DsrC via in-gel digestion, also reported by Gaucher et al.(4).

Molecular mass measurements of intact complexes demonstrated a difference in stoichiometry between the two early eluting dissimilatory sulfite reductase forms: the first fraction contained a hexamer [DsrA]<sub>2</sub>[DsrB]<sub>2</sub>[DsrC]<sub>2</sub>, while the second fraction contained a pentamer [DsrA]<sub>2</sub>[DsvB]<sub>2</sub>[DsvC], both measured within an error of less than 0.1%. The third fraction contained a tetramer [DsrA]<sub>2</sub>[DsvB]<sub>2</sub>. The accuracy of mass measurement allowed for assignment of the number of polypeptide subunits within each complex. However, the exact number and structure of the prosthetic groups involved was not discerned. The products of partial protein complex dissociation (either in solution or in the gas phase) suggested that the [DsrA][DsrB] carrying prosthetic groups formed the core of all three forms of Dsr complex.

#### References

1. Crane and Getzoff (1996) *Curr. Op. Struct. Biol.* 6:744
2. Pierik et al, (1992) *Eur. J. Biochem.* 205:11
3. Marritt and Hagen (1996) *Eur J Biochem*, 238:724
4. Gaucher et al. (2008) *J. Proteome Res.* 7: 2320

GTL

### Microbes Online: An Integrated Portal for Comparative Functional Genomics

Paramvir S. Dehal<sup>1,2\*</sup> (PSDehal@lbl.gov), Jason K. Baumohl,<sup>1,2</sup> Dylan Chivian,<sup>1,2,3</sup> Katherine H. Huang,<sup>1,2</sup> Marcin P. Joachimiak,<sup>1,2</sup> Keith Keller,<sup>1,2</sup> Morgan N. Price,<sup>1,2</sup> and **Adam P. Arkin**<sup>1,2,3,4</sup>

<sup>1</sup>Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov>; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>3</sup>DOE Joint BioEnergy Institute, Emeryville, Calif.; and <sup>4</sup>Dept. of Bioengineering, University of California, Berkeley, Calif.

**Project Goals: MicrobesOnline seeks to integrate functional genomic data with comparative genome sequence to provide novel web-based viewing and analysis tools for gene expression microarray, proteomic, and phenotype microarray data. Selecting an organism or gene of interest in MicrobesOnline leads to information about and data viewers for experiments conducted on that organism and involving that gene or gene product. 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).**

Since 2003, MicrobesOnline (<http://www.microbesonline.org>) has been providing a community resource for comparative and functional genome analysis. The portal includes over 1000 complete genomes of bacteria, archaea and fungi, as well as 1000s of viruses and plasmids. In addition to standard comparative genomic analysis, including gene prediction, sequence homology, domain identification, gene family assignments and functional annotations from E.C. and GO, MicrobesOnline integrates data from functional genomics and places it within a phylogenetic context. Currently, MicrobesOnline has 1000s of microarray based expression experiments from diverse set of organisms ranging from model organisms such as *E. coli* and *S. cerevisiae* to environmental microbes such as *Desulfovibrio vulgaris* and *Shewanella oneidensis*. MicrobesOnline offers a suite of analysis and tools including: a multi-species genome browser, operon and regulon prediction methods and results, a combined gene and species phylogeny browser, phylogenetic profile searches, a gene expression data browser with gene expression profile searches, a gene ontology browser, a workbench for sequence analysis (including sequence motif detection, motif searches, sequence alignment and phylogeny reconstruction), integration with RegTransBase, and capabilities for community annotation of genomes. The next update of MicrobesOnline will contain significant new functionality, including comparative analysis of metagenomic sequence data.

MicrobesOnline seeks to integrate functional genomic data with comparative genome sequence to provide novel web-based viewing and analysis tools for gene expression microarray, proteomic, and phenotype microarray data. Selecting an organism or gene of interest in MicrobesOnline leads to information about and data viewers for experiments conducted on that organism and involving that gene or gene product. It is possible to view microarray data from multiple 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). These new compendium-wide functionalities allow the user 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 MicrobesOnline serves not only to generate insights into the gene expression responses and their regulation in these microorganisms, but also to document experiments, allow contextual access to experimental data, and facilitate the planning of future experiments. MicrobesOnline is actively incorporating publicly available functional genomics data from published research, so as to centralize data on microbial physiology and ecology in a unified comparative functional genomic framework.

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, comparative sequence and phylogenetic data to understand mechanisms

by which microbes survive in uncertain environments while carrying out processes of interest for bioremediation and energy generation. To support this work, VIMSS has developed this web portal (MicrobesOnline), an underlying database, and analyses for comparative functional genomics of bacteria, archaea, fungi and viruses.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

—  
GTL

### Progress in the Development of the RegTransBase Database and the Comparative Analysis System

Michael J. Cipriano,<sup>1</sup> Pavel Novichkov,<sup>1,4</sup> Alexei E. Kazakov,<sup>2</sup> Dmitry Ravcheev,<sup>2</sup> **Adam P. Arkin**<sup>1,3,4</sup> (APArkin@lbl.gov), Mikhail S. Gelfand,<sup>2,5</sup> and Inna Dubchak<sup>1,4,6\*</sup> (ILDubchak@lbl.gov)

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>2</sup>Institute for Information Transmission Problems, RAS, Moscow, Russia; <sup>3</sup>Dept. of Bioengineering, University of California, Berkeley, Calif.; <sup>4</sup>Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov>, Berkeley, Calif.; <sup>5</sup>Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia; and <sup>6</sup>DOE Joint Genome Institute, Walnut Creek, Calif.

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>.

Over 1000 additional articles were annotated last year resulting in the total number of 5118 articles. We specifically focused on annotating the facts of regulation in bioenergy-related bacteria such as: *Clostridia*, *Thermoanaerobacter*, *Geobacillus stearothermophilus*, *Zymomonas*, *Fibrobacter*, *Ruminococcus*, *Prevotella*, *Acetobacter*, *Anaeromyxobacterium*, *Streptomyces*, *Ralstonia*.

Currently, the database describes close to 12000 experiments (30% growth in the last year) in relation to 531 genomes. It contains data on the regulation of ~39000 genes and evidence for ~10000 interactions with ~1130 regulators. We removed redundancy in the list of Effectors (currently the database contains about 630 of them) and turned them into controlled vocabulary.

RegTransBase additionally provides an expertly curated library of 150 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 of the binding site on a published genome, and links to published articles. RegTransBase builds upon these alignments by providing a set of computational modules for the comparative analysis of regulons among related organisms.

The new tool—“advanced browsing” was developed to allow a user to search the data contained in RegTransBase in a step-by-step manner. Different types of classifications, such as taxonomy, effectors, the type of experimental result, a phenotype, and genome relevance allow for creating and applying complex search criteria. We are planning to include additional classifications such as metabolic pathways and types of experimental techniques in this scheme.

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.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

—  
GTL

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

**Terry C. Hazen**<sup>1,2\*</sup> (TCHazen@lbl.gov), Gary Anderson,<sup>1,2</sup> Sharon Borglin,<sup>1,2</sup> Eoin Brodie,<sup>1,2</sup> Steve van Dien,<sup>8</sup> Matthew Fields,<sup>1,7</sup> Julian Fortney,<sup>1,2</sup> Jil Geller,<sup>1,2</sup> E. Hendrickson,<sup>5</sup> Kristina L Hillesland,<sup>1,6</sup> Hoi-Ying Holman,<sup>1,2</sup> J. Leigh,<sup>1,6</sup> T. Lie,<sup>1,6</sup> Janet Jacobsen,<sup>1,2</sup> Dominique Joyner,<sup>1,2</sup> Romy Chakraborty,<sup>1,2</sup> Martin Keller,<sup>1,3</sup> Aindrila Mukhopadhyay,<sup>1,2</sup> Christopher Schadt,<sup>1,3</sup> David Stahl,<sup>1,6</sup> Sergey Stolyar,<sup>1,6</sup> Chris Walker,<sup>1,6</sup> Judy Wall,<sup>1,5</sup> Zamin Yang,<sup>1,3</sup> Huei-che Yen,<sup>1,5</sup> Grant Zane,<sup>1,5</sup> and Jizhong Zhou<sup>1,9</sup>

<sup>1</sup>Virtual Institute of Microbial Stress and Survival; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>4</sup>Diversa, Inc., San Diego, Calif.; <sup>5</sup>University of Missouri, Columbia, Mo.; <sup>6</sup>University of Washington, Seattle, Wash.; <sup>7</sup>Montana State University, Bozeman, Mont.; <sup>8</sup>Genomatica, San Diego, Calif.; and <sup>9</sup>University of Oklahoma, Norman, Okla.

**Project Goals: The goals of this project are: (i) to develop a stable syntrophic microbial consortia in continuous**

flow systems which can be used for physiological and functional genomic studies consisting of *Clostridium cellulolyticum*, *Desulfovibrio vulgaris*, *Geobacter sulfurreducens* and *Methanococcus maripaludis* to simulate community interactions in anaerobic systems, (ii) to study how organisms change their activity on a molecular level in response to other community members in the co-culture by analyzing the gene transcription, metabolite flux and growth, and (iii) to study how these changes in activity and community organization affect the resistance and resilience of microbial communities in response to stress, invasion and other perturbations on a molecular level.

### Field Studies

*Environmental Characterizations.* 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 Hanford 100H as part of our ERSP project we injected 40 lbs of HRC (polylactate) as a slow release electron donor in August 2004. Until March 2008 reducing conditions were maintained, along with undetectable levels of Cr(VI) (Hubbard et al, 2008; Faybishenko et al., 2008). During this time the environment was dominated by sulfate reducers and we were able to detect *Desulfovibrio vulgaris*-like organisms with direct fluorescent antibody. We also isolated a strain of *D. vulgaris* that was only 100 genes different from the type strain (this strain is currently being sequenced by JGI). This year we injected 10 lbs of HRC at the same site to determine if there is a 'memory' response and observed H<sub>2</sub>S production after only 23 h. Once the community stabilizes we will begin push-pull stress tests in the field with NO<sub>3</sub> and monitor functional gene, community structure, and stress responses as compared with previously published models by our group with pure cultures. We are also isolating consortia and determining the dominant community structure to compare with our lab studies. To determine if the stress response in biofilm communities is different than groundwater we are also using "bug traps". Bug traps were filled with Hanford sediment in order to stimulate surface-associated populations in the 100-H subsurface. Work is in progress, and initial results suggest that even under non-stimulated conditions, that more nucleic acids can be extracted from sediment compared to an equivalent amount of groundwater.

*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 & 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 & 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 & 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 syntenuous 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. This year we also submitted *D. vulgaris* Miyazaki, *D. salixigena*, *D. desulfuricans* 27774, *D. aesopensis*, *D. vulgaris* RCH1, *D. fructosivorans*, *D. hanfordensis*, *D. sp.* FW1012B, and *Sulfurospirillum barnesii* strain SES-3 for sequencing at JGI. We are also preparing *D. termitidis* for sequencing at JGI. *Desulfovibrio* FW1012B was isolated from ORFRC groundwater during biostimulation for uranium bioreduction. The isolate can reduce sulfate with lactate, pyruvate, and 1,2-propanediol, reduces chromate, and reduces nitrate without growth. In addition, the genome of *Anaeromyxobacter* fw109-5 has been completed. The *Anaeromyxobacter* strain was isolated from ORFRC sediments, and can reduce iron and heavy metals.

*Pyro-sequencing.* Methods are being developed for parsing environmental pyro-sequencing results for analyses with multivariate techniques. A test set was constructed of 3 samples that represented almost 19,000 SSU rDNA gene fragments.

*Multispecies Syntrophic Consortia.* The goals of this project are: (i) to develop a stable syntrophic microbial consortia in continuous flow systems which can be used for physiological and functional genomic studies consisting of *Clostridium cellulolyticum*, *Desulfovibrio vulgaris*, *Geobacter sulfurreducens* and *Methanococcus maripaludis* to simulate community interactions in anaerobic systems, (ii) to study how organisms change their activity on a molecular level in response to other community members in the co-culture by analyzing the gene transcription, metabolite flux and growth, and (iii) to study how these changes in activity and community organization affect the resistance and resilience

of microbial communities in response to stress, invasion and other perturbations on a molecular level. Stable community configurations and basic interaction structures within the community have recently been demonstrated via QPCR analysis of populations and HPLC/GCMS analysis of the major metabolite flux in the system. A four species microarray configuration is currently being evaluated and used for characterization of stable growth interactions transcriptional patterns in the consortia and comparison to the individual species grown in isolation. We have also found a role for carbon monoxide in the energy metabolism of *Desulfovibrio*. Combined physiological and genetic studies have shown a capacity for both production and consumption of carbon monoxide, and identified a CO sensor (CooA) that is required for CO consumption, most likely via CO-dependent regulation of a carbon monoxide dehydrogenase (CODH). We have also observed a rapid evolutionary improvement in the stability, growth rate, and production (biomass yield) of a simple model community comprised of *D. vulgaris* and *M. maripaludis* after 300 generations of evolution in either a uniform (shaken) or heterogeneous (static) environment. A contribution of both populations to evolutionary improvement was demonstrated. However, the relative contribution of each population to improvement was dependent on growth environment (uniform versus heterogeneous).

#### Acknowledgements

This work was part of the Virtual Ins4tute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

## Microfluidic Tools for Single-Cell Genomic Analysis of Environmental Bacteria

Robert J. Meagher,<sup>1,3</sup> Masood Z. Hadi,<sup>1,3</sup> Yooli K. Light,<sup>1,3</sup> Paramvir Dehal,<sup>2,3</sup> Terry Hazen,<sup>2,3</sup> **Adam Arkin**<sup>2,3</sup> ([aparkin@lbl.gov](mailto:aparkin@lbl.gov)), and Anup K. Singh<sup>1,3\*</sup> ([aksingh@sandia.gov](mailto:aksingh@sandia.gov))

<sup>1</sup>Sandia National Laboratories, Livermore, Calif.;

<sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.;

and <sup>3</sup>Virtual Institute for Microbial Stress and Survival, Berkeley, Calif.

**Project Goals: We are developing high-throughput tools for studying bacteria one cell at a time, allowing us to unravel the complicated dynamics of population, gene expression, and metabolic function in mixed microbial communities. Three microfluidic technologies have been developed to enable these studies: (1) automated 16S rRNA FISH for identification of microbial cells in a mixed sample, (2) photonic force cell sorting for selectively isolating a species of interest (3) encapsulation of individual cells in picoliter-volume droplets, followed by genetic analysis.**

Currently available experimental tools such as microarrays and qRT-PCR for studying the genomes and gene expression in mixed populations of environmental bacteria generally require relatively large amounts of starting material (DNA or RNA), and provide population-averaged data. Although these experimental tools provide valuable insights into microbial communities, the pooling of samples severs the link between the genotype and phenotype of each individual cell. We are developing high-throughput tools for studying bacteria one cell at a time, allowing us to unravel the complicated dynamics of population, gene expression, and metabolic function in mixed microbial communities. Three microfluidic technologies have been developed to enable these studies: (1) automated 16S rRNA FISH for identification of microbial cells in a mixed sample, (2) photonic force cell sorting for selectively isolating a species of interest (3) encapsulation of individual cells in picoliter-volume droplets, followed by genetic analysis. These three technologies can be coupled to one another, allowing, e.g. FISH-based identification of a rare species, followed enrichment of the rare species of interest by photonic force cell sorted, followed by single-cell encapsulation and PCR analysis. The droplet technology in particular allows us to scale down conventional (microliter-volume) assays, such as PCR, into much smaller reaction volumes better suited to the size of an individual microbe. By dramatically reducing the reaction volume, the effective concentration of template is increased, reducing amplification artifacts that often arise in single-cell reactions carried out at a conventional scale. Droplets can be generated rapidly (hundreds per second), with very good uniformity in size (<5% variation in droplet diameter), allowing high throughput experiments to be conducted with much better precision than is possible with conventional emulsion techniques. These technologies are currently under development with simple laboratory strains of *E. coli* and other well-characterized organisms such as *D. vulgaris*, and upon validation will then be translated to studying more complex mixed cultures and environmental samples.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Applications of GeoChip for Analysis of Different Microbial Communities

Joy D. Van Nostrand,<sup>1,2\*</sup> Liyou Wu,<sup>1</sup> Patricia Waldron,<sup>1</sup> Ye Deng,<sup>1</sup> Zhili He,<sup>1,2</sup> Weimin Wu,<sup>3</sup> Sue Carroll,<sup>4</sup> Chris Schadt,<sup>4,2</sup> Anthony Palumbo,<sup>4</sup> Dave Watson,<sup>4</sup> Craig Criddle,<sup>3</sup> Phil Jardine,<sup>4</sup> Terry C. Hazen,<sup>5,2</sup> and Jizhong Zhou<sup>1,2</sup>

<sup>1</sup>University of Oklahoma, Norman, Okla.; <sup>2</sup>VIMSS (Virtual Institute for Microbial Stress and Survival), <http://vimss.lbl.gov>; <sup>3</sup>Stanford University, Stanford, Calif.; <sup>4</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>5</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: Microarray technology provides the opportunity to identify thousands of microbial genes or populations simultaneously. Based on GeoChip 2.0, a new version of GeoChip (GeoChip 3.0) has been developed, which has several new features. First, GeoChip 3.0 contains ~25,000 probes and covers ~47,000 sequences for 292 gene families. Second, the homology of automatically retrieved sequences by key words is verified by HUMMER using seed sequences so that unrelated sequences are removed. Third, a universal standard has been implemented so that data normalization and comparison of different microbial communities can be conducted. Fourth, a genomic standard is used to quantitatively analyze gene abundance.**

Microarray technology provides the opportunity to identify thousands of microbial genes or populations simultaneously. Based on GeoChip 2.0, a new version of GeoChip (GeoChip 3.0) has been developed, which has several new features. First, GeoChip 3.0 contains ~25,000 probes and covers ~47,000 sequences for 292 gene families. Second, the homology of automatically retrieved sequences by key words is verified by HUMMER using seed sequences so that unrelated sequences are removed. Third, a universal standard has been implemented so that data normalization and comparison of different microbial communities can be conducted. Fourth, a genomic standard is used to quantitatively analyze gene abundance. In addition, GeoChip 3.0 includes phylogenetic markers, such as *gyrB*. Finally, a software package has been developed to facilitate management of probe design, data analysis, and future updates. The GeoChip has been used to examine dynamic functional and structural changes in microbial communities from many different environments. Here, three examples of studies utilizing the GeoChip to examine microbial communities at contaminated sites are presented. These 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. Sediments from eleven different sampling wells were evaluated. The results showed that different microbial communities were established in different wells and a high

gene overlap was observed from wells within the same treatment zone. Higher microbial functional gene number, diversity and abundance were observed within the active bioremediation zone. The microbial community structure was highly correlated with the hydraulic flow rate and geochemical conditions of the treatment zone, especially pH, manganese concentration and electron donor level. (2) In a different study of the same system, functional community dynamics were examined during a period of oxidation by nitrate. Diversity and gene number decreased after nitrate exposure and while recovery appeared to begin, the gene numbers were still low even 100 days after nitrate exposure. Principle component analysis (PCA) of detected genes indicated a shift in community structure after nitrate exposure. Nitrate exposure appeared to result in long-term changes to the overall community. (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 NO<sub>3</sub><sup>-</sup> contamination. While diversity of nitrogen fixation genes decreased in NO<sub>3</sub><sup>-</sup> 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 NO<sub>3</sub><sup>-</sup> 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. 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, and it is also a powerful tool to link microbial community structures to ecosystem functioning.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Temporal and Spatial Organization within a Syntrophic Bacterial-Archaeal Biofilm

K. Brileya,<sup>1</sup> C. Walker,<sup>2</sup> S. Stolyar,<sup>2</sup> D.A. Stahl,<sup>2</sup> A. Arkin<sup>3</sup> (AParkin@lbl.gov), T.C. Hazen,<sup>3</sup> and M.W. Fields<sup>1\*</sup> (matthew.fields@erc.montana.edu)

<sup>1</sup>Montana State University, Bozeman, Mont.; <sup>2</sup>University of Washington, Seattle, Wash.; and <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: A goal of VIMSS is to determine the molecular determinants that underlie microbial community function and stability.**

The elucidation of how populations 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 determine the molecular determinants that underlie microbial community function and stability. A syntrophic co-culture of the sulfate-reducing bacterium, *Desulfovibrio vulgaris* Hildenborough, and the methanogenic archaeon, *Methanococcus maripaludis*, was selected as a basal community that can directly and indirectly interact as a biofilm. Planktonic growth conditions, in which cells exist as 'non-adhered cells', rarely represent a true state of growth for the majority of microorganisms under *in situ* conditions, and adherent growth is most likely a universal feature. The roles of biofilms have become increasingly more evident in processes from microbial pathogenesis to waste water to metal corrosion; however, relatively little work has been done on anaerobic biofilms, particularly regarding the structure and behavior of non-pathogenic organisms under environmentally relevant conditions. Microbial communities associated with surfaces may incur protection from stresses such as nutrient-limitation, pH, salts, and heavy metals. In addition, proximity and localization within surface-adhered communities may impact functionality in terms of electron- and hydrogen-metabolism. It was hypothesized that hydrogen transfer would dictate co-culture biofilm formation in the absence of sulfate as terminal electron acceptor for *D. vulgaris* and without addition of hydrogen as electron donor for the methanogen. *M. maripaludis* did not form significant biofilms on a glass surface in batch mono-culture experiments, but *D. vulgaris* did. However, *M. maripaludis* did form a pellicle-like structure in batch, static cultures. A biofilm reactor was developed to co-culture *D. vulgaris* and *M. maripaludis* during syntrophic growth, and spatial and temporal organization was characterized using qPCR, epifluorescent microscopy, field emission electron microscopy, methane production and protein and carbohydrate analysis. During early development, the biofilm initiated as a monolayer of *D. vulgaris* cells, and the mainly *D. vulgaris* biofilm contained extracellular filaments that have been previously described. Soon after the development of the *D. vulgaris* biofilm, *M. maripaludis* cells were observed, and the number of planktonic phase cells declined as the number of biofilm cells increased for both popula-

tions. Over time, the methanogenic biofilm stabilized, and the ratio of *D. vulgaris* to *M. maripaludis* cells was approximately 2.5 and this is a similar ratio observed for cultures entirely populated by planktonic cells. However, at later time points, the planktonic populations had a ratio of approximately 0.2, and this ratio was significantly lower compared to biofilm. Both populations had 1 to 2 log more cells in the biofilm than the planktonic phase. As the methanogenic biofilm developed, extracellular structures continued to be observed. The results suggested that *D. vulgaris* initiated and established a biofilm that then recruited *M. maripaludis*, and the biofilm grew and changed over time as the numbers of both populations increased.

### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Evolution and Stability in a Syntrophic Community

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

University of Washington, Seattle, Wash.

**Project Goals: We compared 20 independently evolved cocultures to cocultures with ancestral *D. vulgaris* and *M. maripaludis*. All but one of the evolved cocultures could grow significantly faster and obtain a much higher yield from the same resource. We tested whether these changes could be attributed to *D. vulgaris*, *M. maripaludis*, or both species by analyzing mixtures of evolved and ancestral populations and found that increases in coculture yield were caused by the combined adaptations of both species.**

Mutualistic relationships between species may often originate from an opportunistic exchange of byproducts produced by each species, but few have addressed how microorganisms evolve in response to these interactions. We used experimental evolution to observe this process in real time in an interaction between the bacterium *Desulfovibrio vulgaris* and the archaeon, *Methanococcus maripaludis*. This mutualism based on the exchange of hydrogen represents an interaction that is fundamental to many microbial communities. After 300 generations of evolution, we compared 20 independently evolved cocultures to cocultures with ancestral *D. vulgaris* and *M. maripaludis*. All but one of the evolved cocultures could grow significantly faster and obtain a much higher yield from the same resource. We tested whether these changes could be attributed to *D. vulgaris*, *M. maripaludis*, or both species by analyzing mixtures of evolved and ancestral populations and found that increases in coculture yield were caused by the combined adaptations of both species. Whether *D. vulgaris*, *M. maripaludis*, or

both contributed to improvements in coculture growth rate depended on the heterogeneity of the evolution environment. Several cocultures had evolved in tubes that were not shaken, so that gradients of hydrogen could form during incubation. Both *D. vulgaris* and *M. maripaludis* that evolved in this environment were capable of enhancing syntrophic growth rate when paired with the ancestor, but *M. maripaludis* that evolved in a constantly shaken homogeneous environment had little if any effect. The evolutionary process leading to these adaptations involved dramatic changes in stability of cocultures over time. After ~30 generations of evolution, cocultures began occasionally entering phases of very slow growth, but eventually these phases stopped, suggesting that one or both species fixed mutations that would stabilize the coculture. Together, these results show that the *D. vulgaris* and *M. maripaludis* strains we studied have a capacity for rapid evolutionary improvements in syntrophic growth, but the process enabling this change depends on the heterogeneity of the environment, and may cause instability in the community.

#### Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

### Development and Analysis of Multispecies Consortia to Study Microbial Community Stress and Survival

L.D. Miller,<sup>1</sup> A. Venkateswaran,<sup>1</sup> J. Mosher,<sup>1</sup> M. Drake,<sup>1</sup> Z.K. Yang,<sup>1</sup> M. Rodriguez,<sup>1</sup> S.D. Brown,<sup>1</sup> T.J. Phelps,<sup>1</sup> M. Podar,<sup>1</sup> A.V. Palumbo,<sup>1</sup> C.W. Schadt,<sup>1</sup> M. Keller,<sup>1</sup> D.C. Joyner,<sup>2</sup> T.C. Hazen,<sup>2</sup> S. Stolyar,<sup>3</sup> K. Hillisland,<sup>3</sup> and D.A. Stahl<sup>3</sup>

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>2</sup>Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>3</sup>Dept. of Civil and Environmental Engineering, University of Washington, Seattle, Wash.

Cultivation of individual microbial species has been at the core of experimental microbiology for more than a century but offers only a glimpse into the metabolism and ecophysiological potential of most microorganisms. Microbial communities, not individual species, control process rates and drive key biogeochemical cycles, including those that determine the transformation of environmental pollutants of concern to the DOE. Thus controlled studies of model consortia comprised of multiple species that mediate such processes are essential for advancing DOE objectives in bioremediation and other applications.

As part of phase two of the Environmental Stress Pathway Project (ESPP) team, the goals of this project are: (i) to develop a stable syntrophic microbial consortia in continu-

ous flow systems which can be used for physiological and functional genomic studies, (ii) to study how organisms change their activity on a molecular level in response to other community members in the co-culture by analyzing the gene transcription, metabolite flux and growth, and (iii) to study how these changes in activity and community organization affect the resistance and resilience of microbial communities in response to stress, invasion and other perturbations on a molecular level.

The project workflow has two major and interrelated components: microbial cultivation and stress experiments and development and application of analytical methods to characterize the consortia. *Clostridium cellulolyticum* was chosen as the basal organism for initial the three & four member syntrophic assemblies as it can ferment cellobiose, producing acetate, lactate, ethanol and hydrogen. The secondary stage in the syntrophic chain is represented by *Desulfovibrio vulgaris* and by *Geobacter sulfurreducens*, which utilize the *C. cellulolyticum*-produced metabolites. *D. vulgaris* and *G. sulfurreducens* are provided with sulfate and fumarate respectively as electron acceptors. Additional studies with a methanogens (*Methanococcus maripaludis* and other species) are also being pursued. Methods for tracking population dynamics such as quantitative PCR and species specific fluorescently labeled antibodies have been developed and have shown that stable assemblages comprised of these species can be achieved. The chemostat instrumentation operates multiple fermenters that are fed from the same medium source. The multiple fermentation setup can be used for providing biological replicates as well as for conducting stress experiments. Methods for quantitative metabolite analysis via HPLC, GC/MS have been successfully developed and the data obtained from analytical methods are consistent with the expected growth and metabolite uptake and depletion patterns for the consortium.

Our initial metabolic analyses of the trimember consortia indicates that growth of *C. cellulolyticum* is carbon limited by the cellobiose concentration. However, its fermentative product acetate remains in abundance despite serving as a carbon source for *D. vulgaris* and a carbon and electron source for *G. sulfurreducens*. *D. vulgaris* likely relies on hydrogen for an electron source and the small amount of lactate from *C. cellulolyticum* and is thereby growth limited by electron donor potential. Growth of *G. sulfurreducens* is limited by abundance of its electron acceptor, fumarate. In experiments with the methanogen (*M. maripaludis*) added to the consortia its populations are inherently less stable and hover at levels nearing the detection limit of the QPCR assay, although methane can often be detected at low levels suggesting biological activity of these organisms. After achieving this stable configuration, we have recently developed a four species microarray for tracking transcriptional activity of these species and comparisons with parallel mass spectrometry-based proteomics. We have begun analyzing data from growth experiment designed to compare stable consortia conditions contrasted with individual species grown alone, and are planning near term experiments to understand resistance and resilience of these communities under various perturbation scenarios on a molecular level.

**Acknowledgements**

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

GTL

## Characterization of Metal-Reducing Communities and Isolates from Uranium-Contaminated Groundwater and Sediments

B. Ramsay,<sup>1,8</sup> C. Hwang,<sup>1,8</sup> S. Carroll,<sup>2</sup> A. Lapidus,<sup>3</sup> J.C. Detter,<sup>3</sup> C. Han,<sup>3</sup> M. Land,<sup>3</sup> L. Hauser,<sup>3</sup> T.C. Hazen,<sup>4,8</sup> A. Arkin<sup>4,8</sup> (AParkin@lbl.gov), A. Belieav,<sup>5</sup> R. Sanford,<sup>6</sup> F. Loeffler,<sup>7</sup> and M.W. Fields<sup>1,8\*</sup> (matthew.fields@erc.montana.edu)

<sup>1</sup>Montana State University, Bozeman, Mont.; <sup>2</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>3</sup>DOE Joint Genome Institute, Walnut Creek, Calif.; <sup>4</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.; <sup>5</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>6</sup>University of Illinois; <sup>7</sup>Georgia Institute of Technology, Atlanta, Ga.; and <sup>8</sup>Virtual Institute for Microbial Stress and Survival

**Project Goals: A goal of VIMSS is to characterize the responses of bacterial populations at multiple levels of resolution in order to understand biochemical capacity at DOE waste sites.**

The elucidation of how metal-reducing bacteria 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 ecosystem function and stability. A goal of VIMSS is to characterize the responses of bacterial populations at multiple levels of resolution in order to understand biochemical capacity at DOE waste sites. Within this context, bacteria with desired functions (e.g., heavy metal reduction) have been isolated from the Oak Ridge Field Research Center (ORFRC). The ORFRC is located within the Y-12 Security Complex near Oak Ridge, TN in the Bear Creek Valley, and the site includes 243-acres of a previously disturbed contaminated area. The subsurface at the FRC contains one of the highest concentration plumes of mobile uranium located in the United States, and contains various levels of nitrate, heavy metal, and organic contamination (<http://www.esd.ornl.gov/orifrc/>).

Recently, biostimulation with ethanol was used to detoxify the contaminated groundwater. The experiment successfully reduced nitrate and uranium levels to a safe-to-humans level during the span of the trial. During the biostimulation, d-*Proteobacteria* were detected to predominate the subsurface groundwater, and sequences indicative of the genera *Desulfovibrio*, *Geobacter*, and *Anaeromyxobacter* were

observed. Two isolates were achieved, *Anaeromyxobacter* fw109-5 and *Desulfovibrio* FW1012B. The *Anaeromyxobacter* genome has been sequenced at JGI, and the *Desulfovibrio* genome is underway.

*Desulfovibrio* FW1012B was isolated from well FW101-2B in the bio-stimulation zone during U(VI) reduction at FRC. The isolate can reduce sulfate and utilize pyruvate, fumarate, maleate, lactate, and 1,2-propanediol. Cr(VI) is reduced, and nitrate is reduced without growth. The isolate can utilize triethylphosphate, metaphosphate, and trimetaphosphate as a phosphorus source. The closest cultivated relative is *Desulfovibrio carbinoliphilus* based upon the SSU rDNA gene sequence, and the closest relative with a completely sequenced genome is *Desulfovibrio fructosovorans*. (that was already said above)

*Anaeromyxobacter* fw109-5 is a mesophilic, iron-reducing bacterium that was isolated from groundwater that had a pH of 6.1 and contained approximately 1.4 mM nitrate and 0.9 μM hexavalent uranium. *Anaeromyxobacter* species are high G+C d-*Proteobacteria* related to the genus *Myxococcus*. Based upon SSU rRNA gene sequences, the closest cultivable relative is *Anaeromyxobacter dehalogenans* 2CP-C with 96.5% sequence identity. The strain fw109-5 grows in the pH range of 4.0 to 9.0, but optimal growth is observed from pH 7.0 to 8.0. To date, known electron donors include acetate, lactate, ethanol, and pyruvate, and electron acceptors include nitrate and iron(III) but not AQDS. Yeast extract and peptone do not support growth, and the organism requires low substrate concentrations for growth (i.e., oligotrophic conditions). Optimal growth occurs under anaerobic conditions, and microaerophilic conditions can be tolerated. The *Anaeromyxobacter* fw109-5 genome is approximately 5.3 Mb in size with 4,336 candidate protein-coding genes. The slow-growing bacterium is predicted to have two SSU *rrn* genes (16S), and almost 30% of the predicted ORFs are classified as conserved hypothetical proteins. A large percentage of estimated ORFs are predicted to be part of a signal transduction pathway with enrichment in serine/ threonine kinase putative proteins. In comparison, fw109-5 had similar numbers of putative two-component and one-component signal transduction proteins as other sulfate- and metal-reducing d-*Proteobacteria*, but fewer compared to *Myxococcus xanthus*. In addition, preliminary data suggest social behavior and sporulation. The genome is predicted to encode a full glycolytic and tricarboxylic acid cycle as well as a pyruvate dehydrogenase complex. Approximately 105 putative proteins are predicted to contain heme-binding sites, with almost half being multi-heme proteins.

**Acknowledgements**

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.



## Changing Patterns of Selection on $\gamma$ -Proteobacteria Revealed by the Ratio of Substitutions in Slow:Fast-Evolving Sites

Jesse Shapiro<sup>1</sup> and Eric Alm<sup>1,2\*</sup> (ejalm@mit.edu)

<sup>1</sup>Massachusetts Institute of Technology, Cambridge, Mass.; and <sup>2</sup>The Virtual Institute for Microbial Stress and Survival, Berkeley, Calif.

**Project Goals: This work is consistent with the DOE stated goal “to gain insights about fundamental biological processes”**

Different microbial species are thought to occupy distinct ecological niches, subjecting each species to unique selective constraints, which may leave a recognizable signal in their genomes. Thus, it may be possible to extract insight into the genetic basis of ecological differences among lineages by identifying unusual patterns of substitutions in orthologous gene or protein sequences. We use the ratio of substitutions in slow versus fast-evolving sites (S:F) to quantify deviations from the typical pattern of selective constraint observed across bacterial lineages. We propose that elevated S:F in one branch (an excess of slow-site substitutions) can indicate a functionally-relevant change, due to either positive selection or relaxed constraint. In a genome-wide comparative study of  $\gamma$ -proteobacterial proteins, we find that cell-surface proteins involved with motility and secretion functions often have high S:F ratios, while information-processing genes do not. Change in evolutionary constraints in some species is evidenced by increased S:F ratios within functionally-related sets of genes (e.g. energy production in *Pseudomonas fluorescens*), while other species apparently evolve mostly by drift (e.g. uniformly elevated S:F across most genes in *Buchnera spp.*). Overall, S:F reveals several species-specific, protein-level changes with potential functional/ecological importance. As microbial genome projects yield more species-rich gene-trees, the S:F ratio will become an increasingly powerful tool for uncovering functional genetic differences among species.

## Natural Diversity and Experimental Evolution of Environmental Stress Tolerance in Marine Bacteria

A.C. Materna,<sup>1\*</sup> S.A. Clarke,<sup>1</sup> C. Cruz,<sup>1</sup> X. Gao,<sup>1</sup> and E.J. Alm<sup>1,2,3</sup> (ejalm@mit.edu)

<sup>1</sup>Massachusetts Institute of Technology, Cambridge, Mass.; <sup>2</sup>The Virtual Institute for Microbial Stress and Survival, Berkeley, Calif.; and <sup>3</sup>The Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Mass.

**Project Goals: This work is consistent with the DOE stated goal “to gain insights about fundamental biological processes”.**

Genome sequencing has revealed extensive genetic variation within bacterial species and among co-existing bacteria. Using marine *Vibrio* strains as a model system, we investigate to what extent observed sequence diversity corresponds to measurable differences in salinity and temperature tolerance phenotypes, two ecologically important factors for this group of organisms. Using directed evolution, we quantify how malleable these phenotypes are with respect to a small number of mutation events. We have designed two-dimensional gradients in 24 cm square dishes containing solid growth medium to monitor temperature and salinity tolerances over a broad range of both factors. Growth patterns indicate the strain-specific minimum and maximum tolerances and interactions between the factors (salinity and temperature). We compared the specific boundaries of growth for multiple strains of *Vibrio splendidus* and *V. alginolyticus*. While the obtained profiles differ in their shape and limits, some consistent features appear. Tolerance to increasing salinities correlated positively with temperature tolerance. However, higher salinity constrained the limits of temperature tolerance, so that the maximum salinity tolerance occurred at intermediate temperatures. Similarly, growth at higher temperatures led to a tradeoff, limiting the range of salinity tolerance. Interestingly, at high salinities, low temperatures tended to suspend growth, leaving viable cells that could be regenerated when the temperature gradient was removed, while higher temperatures led to killing. In addition to comparing related environmental isolates, this method was further applied to study differences between parental and evolved strains. Serial application of  $10^6$  cells/cm<sup>2</sup> to the solid medium gradients enabled selection for spontaneous, more tolerant mutants. Five rounds of this method has repeatably produced mutants that tolerate salinity levels that kill the ancestor. In future work, this integrated ecological and experimental approach will be combined with genome re-sequencing to draw connections between genetic diversity and ecologically relevant phenotypes and tradeoffs.

## The *Shewanella* Federation

GTL

### Constraint-Based Metabolic Modeling Reveals High Maintenance Energy Requirement for Growth of *Shewanella oneidensis* MR-1 Under Elevated O<sub>2</sub> Tensions

Grigoriy E. Pinchuk<sup>1\*</sup> (grigoriy.pinchuk@pnl.gov), Jennifer L. Reed,<sup>2</sup> Oleg V. Geydebrekht,<sup>1</sup> Eric A. Hill,<sup>1</sup> Alexander S. Beliaev,<sup>1</sup> and Jim K. Fredrickson<sup>1</sup>

<sup>1</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and <sup>2</sup>Dept. of Chemical and Biological Engineering, University of Wisconsin, Madison, Wis.

**Project Goals:** The research proposed in this objective is aimed at improving our understanding of the complex nature of energy conversion networks in *Shewanella* including electron acceptor and electron donor utilization pathways. In that regard, we will emphasize providing a system-level conceptual model linking the ETN and the CCM/PCM pathways in MR-1 that will reflect the ecophysiology of *Shewanella* in redox stratified environments.

*Shewanella* species belong to a diverse group of bacteria that live at redox interfaces present in the different environments including water column, sediments and subsurface. The properties of their habitats have shaped *Shewanella* metabolic flexibility allowing these microorganisms to effectively respond to the challenge of rapid changes in nutrient and electron acceptor type and concentration. As part of the Genomics:GTL funded *Shewanella* Federation consortium, which conducts integrated genome-based research of *Shewanella* ecophysiology, we have studied the metabolic response of *S. oneidensis* MR-1 to changes in O<sub>2</sub> concentration during steady-state aerobic growth.

A constraint-based metabolic model of *S. oneidensis* was used to estimate ATP maintenance parameters for MR-1 grown under aerobic conditions on lactate in a chemostat with high (20% air saturation) dissolved oxygen tension (DOT). Using lactate uptake rate measurements and dilution rates as constraints, we calculated the maximum amount of ATP that could be hydrolyzed ( $V_{ATP}^M$ , mmol ATP hydrolyzed / g of dry weight biomass / hr). A linear relationship between cellular growth rate and maximal ATP hydrolyzation was found:

$$V_{ATP}^M \text{ (mmol/gDW/hr)} = 1.2 + 225 \cdot \mu \quad (\text{Eq. 1})$$

Based on these values, the non-growth associated ATP maintenance (NGAM) term was 1.2 (mmol/gDW/hr), which is similar to values for other organisms: *Escherichia coli* (7.6), *Geobacter sulfurreducens* (0.45), and *Bacillus subtilis* (9). However, the growth associated ATP maintenance (GAM)

term (225 mmol/gDW) was significantly higher than has been reported for other bacteria: *E. coli* (~46), *G. sulfurreducens* (~47), and *B. subtilis* (~104).

These ATP estimates were consistent for six different dilution rates when the DOT in the reactor was high (note that only four dilution rates were used in fitting the  $V_{ATP}^M$  equation). We also calculated the  $V_{ATP}^M$  for conditions under O<sub>2</sub> limitation and low DOT, and found that the maximal amount of ATP that was hydrolyzed under these conditions was much lower (about 3-fold) than would be estimated using the equation 1 above. Several hypotheses were generated to explain the high ATP hydrolysis rates observed under elevated O<sub>2</sub> concentrations: (i) oxidative damage of redox proteins enhances their intracellular turnover rate; (ii) *S. oneidensis* MR-1 forms aggregates containing a major extracellular polymeric substance (EPS) component in the presence of mmolar concentration range of Ca<sup>2+</sup> ions under elevated O<sub>2</sub> levels. Cell aggregation is thought to be a defense mechanism against oxidative stress. DNA, carbohydrates, and proteins are known to be an important part of EPS in these aggregates. We grew bacteria in a medium that contained trace CaCl<sub>2</sub>, therefore no aggregation was observed although presumably EPS was still produced. We hypothesize that the extracellular polymers produced under highly aerobic conditions where hydrolyzed and then recycled by growing cells thus generating a futile cycle reflected in elevated ATP spending; (iii) oxidative stress caused by high DOT triggered cell lysis possibly as a result of lysogenic conversion since *S. oneidensis* MR-1 chromosome contains DNA encoding for 3 prophages. Experimental assessment of these hypotheses is underway, however first TEM images taken for high O<sub>2</sub> cultures showed significant amounts of viral particles present in *S. oneidensis* growth medium.

This work represents an important step toward understanding of *Shewanella* species ecophysiology using constraint-based analysis as a key tool for investigating energy metabolism.

GTL

### Phylogenetic Footprinting in the *Shewanellae*

Lee Ann McCue<sup>1\*</sup> (leeann.mccue@pnl.gov), William A. Thompson,<sup>2</sup> Lee A. Newberg,<sup>3,4</sup> and Charles E. Lawrence<sup>2</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>2</sup>Center for Computational Molecular Biology and the Division of Applied Mathematics, Brown University, Providence, R.I.; <sup>3</sup>The Wadsworth Center, New York State Dept. of Health, Albany, N.Y.; and <sup>4</sup>Dept. of Computer Science, Rensselaer Polytechnic Institute, Troy, N.Y.

**Project Goals:** The major goal of this project is the development of the computational bioinformatics tools neces-

## sary for study of transcription regulation in prokaryotes by comparative genomics.

Species of the genus *Shewanella* have diverse metabolic capabilities, including their ability to reduce various terminal electron acceptors, leading to considerable interest in their potential for remediation of contaminated environments and use in microbial fuel cells. This interest has spurred the genome sequencing and concomitant comparative genome analysis of over 20 species of *Shewanella*, each with distinct physiology and isolated from various natural habitats (Fredrickson, et al., 2008). As with other comparative genomics studies, the study the transcription regulatory mechanisms of the *Shewanellae* has been facilitated by the availability of these genome data, while simultaneously being confounded by the close phylogenetic relationship of these species. Specifically, while the identification of transcription factor binding sites and regulons can be predicted by analysis of orthologous promoter data (*i.e.*, phylogenetic footprinting), the recent speciation of closely related genomes results in correlation in the sequence data, which confounds the detection of functionally conserved sequence motifs by conventional motif-finding methods.

To address this challenge, we have developed a version of the Gibbs recursive sampler that incorporates phylogeny of the input sequences through the use of an evolutionary model, and calculates an ensemble centroid motif solution (Newberg, et al., 2007); these extensions yield an improved ability to avoid false positive motif prediction in phylogenetically correlated data. We have applied these advanced features of the phylogenetic Gibbs sampler to predict regulatory sites in orthologous data from 17 *Shewanellae*. Furthermore, because many of the predicted sites are likely bound by a common transcription factor, motifs with similar patterns were clustered to infer sets of co-regulated genes (*i.e.*, regulons).

## References

1. Fredrickson, J. K., Romine, M. F., Beliaev, A. S., Auchtung, J. M., Driscoll, M. E., Gardner, T. S., et al. (2008). Towards environmental systems biology of *Shewanella*. *Nat Rev Microbiol*, 6(8), 592-603.
2. Newberg, L. A., Thompson, W. A., Conlan, S., Smith, T. M., McCue, L. A., & Lawrence, C. E. (2007). A phylogenetic Gibbs sampler that yields centroid solutions for cis-regulatory site prediction. *Bioinformatics*, 23(14), 1718-1727.

## Constraint-Based Modeling of Metabolism in *Shewanella oneidensis* MR-1

Jennifer L. Reed<sup>1\*</sup> (reed@engr.wisc.edu), Sharon J. Wiback,<sup>2</sup> Iman Famili,<sup>2</sup> Jessica De Ingeniis,<sup>3</sup> and Grigoriy Pinchuk<sup>4</sup>

<sup>1</sup>Dept. of Chemical and Biological Engineering, University of Wisconsin, Madison, Wis.; <sup>2</sup>Genomatica, Inc., San Diego, Calif.; <sup>3</sup>Burnham Institute for Medical Research, La Jolla, Calif.; and <sup>4</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.

**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 760 reactions, 780 genes, and 623 metabolites. A biomass equation specific for *S. oneidensis* has been developed based on published data, and phospholipid and amino acid composition measurements taken under aerobic conditions.

The reconstruction was used to build a constraint-based 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 four classes of electron acceptors, with differing biomass yields (g D.W. produced per mmol electron acceptor consumed). Gene deletion simulations under different environmental conditions with various carbon sources and electron acceptors identified a large group of genes were never essential (540 out of 780), and a smaller fraction that were always essential (200 out of 78) for growth.

The developed model is 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 part 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.

## Acknowledgements

The co-authors wish to thank all members of the *Shewanella* Federation, especially Margaret Romine, Gretta Serres, and Andrei

Osterman for their work on the genome annotations for MR-1 and other *Shewanella*, which was critical for the development of the metabolic model.

GTL

## Progress in Identification of the ‘Mobilome’ Associated with 21 Sequenced *Shewanella* sp.

**Margaret Romine\*** (Margie.romine@pnl.gov)

Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals: The *Shewanella* Federation project has four overarching goals: 1) elucidate the environmentally-controlled signal transduction and transcriptional regulatory systems crucial for the ecological success of *Shewanella*, 2) characterize the electron transfer networks and central metabolic pathways involved in energy conversion in *Shewanella*, 3) develop an integrated model of *Shewanella* metabolism, sensing, & regulation, and 4) develop an understanding of *Shewanella* population genomics: understanding evolution, ecophysiology and speciation. In this presentation we will highlight our initial findings on the contribution of laterally acquired DNA to the evolution of this species.**

Mobile elements play an important role in the evolution of microbial genomes through activities such as horizontal gene transfer, gene disruption, gene expression modulation, and recombination. There are currently 19 complete and 4 partial *Shewanella* genome sequenced derived from strains that vary considerably in phylogenetic type, environmental origin, and culture conditions (e.g., temperature, salinity, carbon and energy sources) that will support maximal growth rates. These genomes are rich in mobile genetic elements including insertion sequences, transposons, bacteriophage, plasmids, miniature inverted-repeat transposable elements (MITEs), group II introns, integrative conjugative elements (ICE), and mobile genomic islands. This combined richness in mobile elements and broad diversity in strain type that has been sequenced provides an excellent resource for studying the evolutionary events that have enabled members of the Genus to inhabit such a broad variety of niches.

Using comparative sequence analysis the precise termini of many of the mobile elements and consequently to predict integration site specificity and to establish what types of functions have been acquired by different *Shewanella* through lateral transfer. Genome sizes across this group vary from 4.3 Mb (*S. amazonensis*) to 5.9 Mb (*S. woodyii*) and predicted to encode from 3668 to 4941 CDS, respectively. So far, up to 15% of the total predicted protein-encoding genes within a *Shewanella* genome are predicted to be encoded by mobile elements. However, we estimate that the number will be even larger as additional elements are discovered and mapped, including those that are more difficult to delineate due to subsequent evolutionary events that have

resulted in attrition or accretion of the originally mobilized element.

Insertion elements are found in all of the genomes and frequently occur within other mobile elements, suggesting that many of the laterally acquired functions do not confer selective advantage to the host and hence their functionality is gradually lost over time. This is not always the case as is exemplified by the large mobile genomic islands that are devoid of IS elements and confer arsenate detoxification and respiratory functionality to *Shewanella* sp. ANA-3 and nitrate assimilatory capability to *S. denitrificans*. The occurrence of genes encoding host-restriction modification enzymes on many of these mobile elements suggest that they also play a significant role in controlling high frequency uptake of DNA from other microbial hosts. Further, the identification of other species that encode genes with high identity to those found in mobile elements of *Shewanella* provide clues as to the types of organisms with which they previously formed close associations with in natural environments and the evolutionary history that resulted in acquisition of new traits or loss of previously encoded ones. For example, the CyaB adenylate cyclase that is encoded in *S. frigidimarina* has between 55 and 64% identity to bi-directional best hits in all of the other sequenced *Shewanella* except *S. denitrificans*, *S. amazonensis*, and *S. woodyii*. However, it possesses slightly better (67–69% identity) similarity to proteins encoded by *Marinomonas*, *Vibro*, and *Allivibrio* and is encoded by mobile elements in both *S. frigidimarina* and *Marinomonas*, while all the remaining *Shewanella* *cyaB* genes do not appear to be encoded by mobile elements.

This finding suggests an independent origin for the *S. frigidimarina* *cyaB* gene relative to those found in the other sequenced *Shewanella*. The same is true of the NADH-ubiquinone oxidoreductase in *S. oneidensis*. In this case there is currently no clear evidence that it is encoded by a mobile element. However, the absence of genes encoding this function in all of the sequenced *Shewanella* and the observation that they encode proteins with greater than 90% identity to those encoded by *Aeromonas hydrophila* suggests that they share an evolutionary origin and there is a high likelihood that members of these Genera interact with each other or a common partner in natural environments.

We anticipate that as additional genome sequence becomes available from projects that explore broader phylogenetic diversity (e.g. JGI's GEBA project and various metagenome projects) than is currently available, it will be possible to more accurately trace the evolutionary history that resulted in diversification of the *Shewanella* group.

### Acknowledgements:

This work was facilitated by use of the Ortholog editor tool in the *Shewanella* Knowledgebase (funded by DOE under the *Shewanella* federation project led by Ed Uberbacher) and the use of neighborhood analysis resources in the JGI Integrated Microbial Genomes (IMG) system and the Fellowship for Interpretation of Genomes SEED system. I would also like to acknowledge LeeAnn McCue for identification of the draft ortholog pairs that were uploaded in the Ortholog Editor, Tatiana Karpinets, Guru Kora, and Denise Schmoyer for development of the Ortholog Editor, and Gretta Serres and Kostas Konstantinidis for identification of strain-specific genes in the first 10 sequenced *Shewanella* genomes.

## Adenylate Cyclases and Anaerobic Respiration in *Shewanella oneidensis* MR-1

M. Charania<sup>1\*</sup> (charania@uwm.edu), E. Biddle,<sup>1</sup> Y. Zhang,<sup>2</sup> G. Pinchuk,<sup>2</sup> A. Beliaev,<sup>2</sup> and D. Saffarini<sup>1</sup> (daads@uwm.edu)

<sup>1</sup>Dept. of Biological Sciences, University of Wisconsin, Milwaukee, Wis.; and <sup>2</sup>Pacific Northwest National Laboratory, Richland, Wash.

**Project Goals:** One of our major goals is to identify the networks that are involved in the regulation of anaerobic respiration in *Shewanella oneidensis* MR-1. We are also interested in identifying the enzymes that are involved in anaerobic respiration of sulfur compounds, and the mechanisms that regulate their expression.

Anaerobic respiration in the metal reducing bacterium *Shewanella oneidensis* MR-1 is regulated by cAMP and its receptor protein CRP. Three *S. oneidensis* proteins, CyaA, CyaB, and CyaC, were shown to have adenylate cyclase activity. Chromosomal deletions of either *cyaA* or *cyaB* did not affect anaerobic respiration with fumarate, DMSO, or Fe(III), whereas deletion of *cyaC* caused deficiencies in respiration with DMSO and Fe(III), but not with fumarate. A double mutant that lacks *cyaA* and *cyaC* was unable to grow anaerobically with the same electron acceptors and exhibited a phenotype similar to that of a *crp* mutant. Microarray analysis of gene expression in the *crp* and *cyaC* mutants indicated that both are involved in the regulation of numerous genes and include the DMSO, fumarate, and Fe(III) reductase genes. Additionally, several genes were differentially expressed in the *cyaC* mutant, but not in the *crp* mutant. Our results indicate that CyaC plays a major role in regulating anaerobic respiration, and may contribute to additional signaling pathways independent of CRP.

In addition to terminal reductase genes, CRP appears to regulate the expression of genes that encode two-component system transduction proteins. These include SO4155 and SO4157. Deletion of these genes in *S. oneidensis* MR-1 led to deficiencies in sulfur, thiosulfate, and tetrathionate reduction. Our results indicate that this deficiency may be due to lack of *phsABC* expression that was predicted to encode the thiosulfate reductase. Mutagenesis of this operon led to significant loss of thiosulfate, tetrathionate, and polysulfide reduction. Thus, it appears that PhsABC is the major enzyme used for the reduction of several sulfur compounds in *S. oneidensis*. This is in contrast to other bacteria, such as *Salmonella typhimurium*, that use different enzymes for the reduction of sulfur compounds.

## Experimental and Computational Analysis of Growth-Phase Dependent Transcriptional Programs in *Shewanella oneidensis*

Qasim K. Beg<sup>1\*</sup> (qasimbeg@bu.edu), Niels Klitgord,<sup>2</sup> Mattia Zampieri,<sup>2,3</sup> Timothy S. Gardner,<sup>1,4</sup> and Daniel Segre<sup>1,2,5</sup> (dsegre@bu.edu)

Depts. of <sup>1</sup>Biomedical Engineering, <sup>2</sup>Bioinformatics Program, and <sup>3</sup>Biology, Boston University, Boston, Mass.; <sup>3</sup>International School for Advanced Studies, Trieste, Italy; and <sup>4</sup>Amyris, Emeryville, Calif.

**Project Goals:** This project is part of the *Shewanella* Federation research activity. As such, it contributes to the overall goal of applying genomic tools to better understanding of the eco-physiology of the respiratory-versatile members of this important genus. Specifically, in our work we expand and improve our knowledge of the physiology of *Shewanella oneidensis* MR-1 by combining measurements of gene expression in time-course experiments under different growth conditions, with computational analysis and flux balance modeling.

In order to study the complex regulatory network of *Shewanella oneidensis* MR-1 and to understand its unique respiratory and metal-reducing abilities we integrate bio-reactor experiments under defined environmental/growth conditions with computational analyses of gene expression clusters and metabolic fluxes.

Understanding the complex nature of bacterial responses to changing environmental conditions requires gaining insights into the global regulatory processes at various phases of bacterial growth. The growth-phase related gene expression changes in *E. coli* have been studied in the past; however, not much is known about physiological and transcriptional changes in *S. oneidensis* when cells pass through exponential, stationary, and transition phases. To examine if transcriptional signatures can help understand these changes in cellular organization, we grew *S. oneidensis* MR-1 in both complex (LB) and minimal medium (M4-Lactate). Time-course samples from various phases of *S. oneidensis* growth were subjected to microarray analysis (using *Shewanella* arrays from Affymetrix) for temporal gene expression study along a time-course.

We processed the probe intensities of the microarrays in each data set using standard quality control and normalization methods, and constructed a temporal expression profile for each probe set. To determine the functional modules in each data set, we constructed clusters of probesets by applying the k-means algorithm using the Pearson correlation metric as a distance between expression profiles. We then looked for statistical enrichment of Gene Ontology functions for each cluster. This statistical analysis revealed several clusters showing functional enrichment of specific gene classes during various phases of growth of the bacterium.

We observed very specific gene expression patterns in the generated clusters during various phases of growth (from early- to mid- to late-log, and stationary phase) in both growth media. These patterns include the up-regulation of prophage and flagella/motility related genes in the complex medium (during exponential phase) and minimal medium (upon entry into stationary phase), respectively. We also observed significant up- and down-regulation of energy metabolism, amino acids biosynthesis, fatty acid and phospholipids metabolism genes, as well as stress response and starvation genes. Finally, we show how additional insight into the regulatory program of *S. oneidensis* can be gained by studying the transcriptional organization of genes along the chromosome.

GTL

## Metabolic Optimality and Trade-Offs Under Combinatorial Genetic and Nutrient Modifications

David K. Byrne<sup>1\*</sup> (Dbyrne@Bu.Edu), Timothy S. Gardner,<sup>2,4</sup> and Daniel Segre<sup>1,2,3</sup>

<sup>1</sup>Graduate Program in Bioinformatics, <sup>2</sup>Dept. of Biomedical Engineering, and <sup>3</sup>Dept. of Biology, Boston University, Boston, Mass.; and <sup>4</sup>Amyris, Emeryville, Calif.

<http://prelude.bu.edu>

**Project Goals: To optimize the metabolic capabilities of *Shewanella oneidensis* and *Escherichia coli*.**

Metabolic engineering in microbial hosts for the production of renewable chemicals and energy sources has received considerable attention in recent years. Lactic acid, biohydrogen, and biodiesel are representative examples of renewable chemical commodities that can be produced using biodegradable and sustainable compounds. Yet the production output and efficiency of these commodities still fall significantly short of theoretical limits and are, in some cases, insufficient for practical implementation. Since microbial metabolism is the primary cellular mechanism by which energy is generated and distributed, it is of interest to understand how metabolic pathways may be modulated to increase the proportion of energy diverted for industrial use.

A computational framework that uses models of microbial metabolism has been developed to design genetic and nutrient programs that optimize industrial output. Multiple engineering objectives are considered: maximization of productivity, yield, and/or purity, minimization of economic cost, and combinations thereof. Associated trade-offs and pareto-optimal sets are determined for different formulations. To search for optimal solutions using linear objectives, the strong duality theorem from linear programming is used to formulate a bilevel optimization procedure. For nonlinear objectives, combinatorial and genetic algorithm approaches are implemented.

Optimal predictions indicate that significant improvements in *Escherichia coli* lactic acid synthesis and *Shewanella oneidensis* reducing power and biohydrogen generation are attainable. Thus nutrient and genetic reengineering may be used to synergistically improve bioenergetic output for sustainable chemical commodity production.

GTL

## Comparative Analyses Across an Evolutionary Gradient Within the *Shewanella* Genus

Margrethe H. Serres<sup>1\*</sup> (mserres@mbl.edu), Kostas Konstantinidis,<sup>2</sup> Margaret F. Romine,<sup>3</sup> Jorge L.M. Rodrigues,<sup>4</sup> Anna Y. Obraztsova,<sup>5</sup> Mary S. Lipton,<sup>3</sup> LeeAnn McCue,<sup>3</sup> James K. Fredrickson,<sup>3</sup> and James M. Tiedje<sup>6</sup>

<sup>1</sup>Marine Biological Laboratory, Woods Hole, Mass.; <sup>2</sup>Georgia Institute of Technology, Atlanta, Ga.; <sup>3</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>4</sup>University of Texas, Arlington, Tex.; <sup>5</sup>University of Southern California, Los Angeles, Calif.; and <sup>6</sup>Center for Microbial Ecology, Michigan State University, East Lansing, Mich.

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

*Shewanella* is an environmentally important bacterial genus with a versatile electron accepting capability as well as the ability to degrade a broad range of carbon compounds. Members of the genus have been isolated from diverse geographic locations and habitats such as marine and fresh water columns, sediments, and subsurface environments where they appear to thrive in the redox interfaces. Genome analyses have revealed an expansion of proteins involved in sensing and responding to the environment, including chemotaxis receptors, two-component regulators, outer membrane proteins/receptors, and proteins for synthesis or degradation of cyclic nucleotides (signaling molecules). Some of these genome changes are species specific and may reflect ecological specialization or the speciation process.

Genome sequences have been obtained for over 20 members of the *Shewanella* genus. Strains were selected for sequencing based on their phylogenetic relatedness and include members with similarity at the sub-species level (*Shewanella* sp. MR-4 and MR-7; *S. putrefaciens* W3-18-1 and CN-32) as well as species that are either closely or distantly related. Together the selected genomes represent organisms within a bacterial genus but along an evolutionary gradient of relatedness. This dataset therefore forms a unique framework for studying evolution of microbial genomes.

In this work ten of the *Shewanella* strains have been analyzed for common (core) features representing func-

tions shared among members of this genus. In addition differences in genomic, proteomic and growth patterns have been examined in order to link genotypic differences to phenotypic changes and to speciation. The analyses revealed that genotypic and phenotypic similarities among the organisms could be predicted from their evolutionary relatedness despite evidence of extensive horizontal gene transfer. However, the power of the predictions did depend on the degree of ecological specialization of the organisms studied. Making use of the genetic gradient formed by the ten genomes, we could distinguish ecological effects from evolutionary divergence and rank cellular functions in terms of rates of evolution.

We also detected that differences in whole-cell protein expression patterns for organisms grown under identical conditions were larger than the differences detected at the orfome level. This suggests that similarity in gene regulation and expression should be included as a parameter when species are described. Overall our results constitute information that may be applied towards understanding bacterial species and genera at a system level. We believe these findings can be transferred to many other ecologically versatile bacteria that are prevalent in important habitats on Earth.

GTL

## Evolution of Signal Transduction in a Bacterial Genus

**Harold Shanafi**\* (hshanafi@utk.edu), Luke E. Ulrich, and Igor B. Zhulin

National Institute for Computational Sciences, University of Tennessee, Oak Ridge National Laboratory, Oak Ridge, Tenn.

The recent completion of sequencing of multiple genomes in the *Shewanella* genus provides a unique opportunity to study evolution at a much finer scale than previously possible. Using a bi-directional best BLAST hit approach at the protein domain rather than traditional whole-protein sequence level we analyzed the evolutionary relationships of proteins predicted to be involved in signal transduction. Based on these relationships, we have determined a core set of 99 proteins across the first 11 sequenced *Shewanella* genomes that were highly conserved in both domain architecture and protein sequence. The core included one of the several chemotaxis systems found in *Shewanella* and several two-component regulatory systems. A large group of orthologous signal transduction proteins across multiple genomes showed some primary sequence drift and were classified as “significant similarity”, and finally there were several unique signal transduction proteins in each organism. We also quantified a recent disproportionate loss of signal transduction genes in *Shewanella denitrificans* OS217 above and beyond the overall reduction in that organism’s genome size, and an enrichment of signal transduction genes in *Shewanella amazonensis* SB2B. Possible relationships of the

observed changes with the metabolism and environment are discussed.

GTL

## Investigating Environmental Specialization in a Population of *Shewanella baltica* Strains

Jie Deng,<sup>1</sup> Jennifer M. Auchtung<sup>1\*</sup> (blackjen@msu.edu), Konstantinos Konstantinidis,<sup>2</sup> and **James M. Tiedje**<sup>1</sup> (tiedje@msu.edu)

<sup>1</sup>Center for Microbial Ecology, Michigan State University, East Lansing, Mich.; and <sup>2</sup>School of Civil and Environmental Engineering and School of Biology, Georgia Institute of Technology, Atlanta, Ga.

**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 the environmentally-versatile members of the important genus.**

Studying how bacterial strains diverge and how divergence correlates with specialization to new environmental niches is important for understanding the dynamics of environmental communities, which often contain populations of closely related strains, and for understanding the potential fate of genetically engineered microbes in the environment. We are studying these questions using a collection of *Shewanella* strains isolated from different regions within the water column of the Baltic Sea. *Shewanella* species have been intensively studied due to their extensive respiratory versatility and their potential role in bioremediation and microbial power generation. Our analysis centers on a collection of *S. baltica* strains previously shown to occupy characteristic positions within the water column. Physical and chemical analysis of the water column performed at the time of strain isolation revealed the presence of a stable redox gradient and provided additional insights into what environmental variables might be important for specialization. Therefore, these strains were ideal candidates for studying strain divergence and environmental specialization.

We used multi-locus sequence typing to analyze phylogeny of 36 strains and examine how this correlates with position within the water column. We identified 8 distinct clades. One clade showed good correlation with depth of isolation; it contained 86% of strains isolated from the anoxic region of the water column. Strains isolated from other regions of the water column were distributed throughout the other clades, suggesting that if specialization to these other depths had occurred, it had likely not occurred at the genome level. We used comparative genomic hybridization (CGH) to provide additional detail of the genetic diversity among these 36 *S. baltica* strains. Genome sequences from four *S. baltica* strains, isolated from three different depths within the water column and previously sequenced through the DOE-GTL program, were used to design the microarrays. Preliminary analysis of this data revealed that when overall similarities

of the genomes were analyzed using a presence/absence matrix, similar phylogenies were produced to those observed through MLST. However, we also identified a subset of genes that were shared primarily by strains isolated from the anoxic region of the water column. Although the majority of these genes encoded proteins with hypothetical or predicted mobile element functions, one cluster encoded genes predicted to be involved in extracellular polysaccharide biosynthesis. Transcriptional profiling revealed that these putative polysaccharide biosynthetic genes were expressed by one of the strains isolated from the anoxic region, providing additional evidence of the potential importance of these genes. Further characterization of these genes, as well as studies comparing differences in gene expression and physiology among these isolates are ongoing.

## Characterization of C-type Cytochromes and Their Role in Anaerobic Respiration in *Shewanella oneidensis* and *S. putrefaciens*

Haichun Gao,<sup>1</sup> Jingrong Chen,<sup>1</sup> Yili Liang,<sup>1\*</sup> Soumitra Barua,<sup>1,2</sup> Dongru Qiu,<sup>1</sup> Zhili He<sup>1</sup> (zhili.he@ou.edu), Alexander S. Beliaev,<sup>3</sup> Margaret F. Romine,<sup>3</sup> Samantha Reed,<sup>3</sup> Dave Culley,<sup>3</sup> David Kennedy,<sup>3</sup> Yunfeng Yang,<sup>2</sup> James M. Tiedje,<sup>4</sup> James K. Fredrickson,<sup>3</sup> Kenneth H. Nealson,<sup>5</sup> and **Jizhong Zhou**<sup>1,2</sup> (jzhou@ou.edu)

<sup>1</sup>Institute for Environmental Genomics and Dept. of Botany and Microbiology, University of Oklahoma, Norman, Okla.; <sup>2</sup>Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>3</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>4</sup>Center for Microbial Ecology, Michigan State University, East Lansing, Mich.; and <sup>5</sup>Dept. of Earth Sciences, University of Southern California, Los Angeles, Calif.

**Project Goals:** This project is focused applying genome-based experimental and modeling approaches to understanding the ecophysiology of *Shewanella* via four overarching and integrated objectives: 1) Elucidate the environmentally-controlled signal transduction and transcriptional regulatory systems crucial for the ecological success of *Shewanella*; 2) Characterize the electron transfer networks and central metabolic pathways involved in energy conversion in *Shewanella*; 3) Develop an integrated model of *Shewanella* metabolism, sensing, & regulation; and 4) Develop an understanding of *Shewanella* population genomics: understanding evolution, ecophysiology and speciation.

*Shewanella* strains are renowned for their ability to utilize a wide range of electron acceptors for respiration, which is due to a large number of c-type cytochromes in their genome. The arsenal of c-type cytochromes is also highly diversified across the 21 sequenced *Shewanella* genomes and only twelve out of the 41 c-type cytochrome of *S. oneidensis* MR-1 are present in all other sequenced strains. Only a

few c-type cytochromes have been characterized so far. To discern the functions of unidentified c-type cytochrome genes in bacterial energy metabolisms, we generated 37 single mutants with in-frame deletion of each individual cytochrome gene in MR-1. Reduction of a variety of electron acceptors was measured and the relative fitness was calculated for these mutants based on competition assays. It was revealed that SO0610, SO1777, SO2361, SO2363, and SO4360 were important under aerobic growth conditions, and that most c-type cytochromes play a more important role in anaerobiosis. The *petC* gene appeared to be important to both aerobiosis and anaerobiosis. Our results regarding functions of CymA and MtrC are consistent with previous findings. We also assayed the biofilm formation of these mutants and it was shown that SO4666 might be important for pellicle formation.

GTL

*S. putrefaciens* W3-18-1 lacks orthologues for the secondary metal reductase and accessory proteins (MtrFED) of *S. oneidensis* MR-1. Sputw3181\_2446 encodes a decaheme c-cytochrome, orthologous to outer membrane primary metal reductase OmcB of MR-1 (60% similarity) while another reductase similar to OmcA in MR-1 was also found in W3-18-1. Sputw3181\_2445 encodes an 11-heme c-type cytochrome OmcE, which only shares 40% similarity with the decaheme cytochrome OmcA. Single and double in-frame deletion mutants of *omcB* and *omcE* were generated for functional characterization of *omcE* and metal reduction in W3-18-1. Reduction of solid-phase Fe(III) and soluble Fe(III) in *S. putrefaciens* W3-18-1 was mainly dependent on OmcB under anaerobic conditions (with 50 mM lactate as electron donors and Fe<sub>2</sub>O<sub>3</sub>, α-FeO(OH), β-FeO(OH) and ferric citrate as electron acceptors. W3-18-1 catalyzed a faster reduction of α-FeO(OH) as compared to MR-1, suggesting that other genes might be involved in iron(III) reduction in W3-18-1. As previously observed in MR-1, the deletion of both OMCs (OmcE & OmcB) led to a severe deficiency in reduction of solid-phase Fe(III) in W3-18-1 and the defectiveness was more remarkable in the reduction of soluble iron. The *omcB* and *omcE* genes of W3-18-1 had been expressed with the pBAD vector in *E. coli*. Heme staining assays also demonstrated that the disappearance of specific protein bands in the SDS-PAGE gels were consistent with *omcB* and *omcE* deletion in three mutant samples. These results suggest that *omcE* and *omcB* are actually expressed as cytochrome proteins and could play a central role in metal reduction in *S. putrefaciens* W3-18-1.



GTL

## The NapC- and CymA-Dependent Nitrate Reduction in *Shewanella oneidensis* MR-1 and *S. putrefaciens* W3-18-1

Dongru Qiu<sup>1\*</sup> (qiu@ou.edu), Haichun Gao,<sup>1</sup> Yili Liang,<sup>1</sup> Jingrong Chen,<sup>1</sup> Zhili He,<sup>1</sup> Margaret F. Romine,<sup>2</sup> Kenneth H. Nealson,<sup>4</sup> James M. Tiedje,<sup>3</sup> James K. Fredrickson,<sup>2</sup> and **Jizhong Zhou<sup>1</sup>** (jzhou@ou.edu)

<sup>1</sup>Institute for Environmental Genomics and Dept. of Botany and Microbiology, University of Oklahoma, Norman, Okla.; <sup>2</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>3</sup>Center for Microbial Ecology, Michigan State University, East Lansing, Mich.; and <sup>4</sup>Dept. of Earth Sciences, University of Southern California, Los Angeles, Calif.

**Project Goals: This project is focused applying genome-based experimental and modeling approaches to understanding the ecophysiology of *Shewanella* via four overarching and integrated objectives: 1) Elucidate the environmentally-controlled signal transduction and transcriptional regulatory systems crucial for the ecological success of *Shewanella*; 2) Characterize the electron transfer networks and central metabolic pathways involved in energy conversion in *Shewanella*; 3) Develop an integrated model of *Shewanella* metabolism, sensing, & regulation; 4) Develop an understanding of *Shewanella* population genomics: understanding evolution, ecophysiology and speciation.**

Nitrate respiration systems are highly diverse among *Shewanella* species. Bioinformatics analyses revealed three types of nitrate reduction systems in *Shewanella* genomes. The well studied *S. oneidensis* MR-1 harbors only the CymA-dependent nitrate reductase and the NapGH ubiquinol oxidase. Most *Shewanella* species including *S. putrefaciens* W3-18-1 have both CymA- and NapC-dependent nitrate reductases, as well as the NapGH ubiquinol oxidase. The *S. baltica* strains have both the CymA- and NapC-dependent nitrate reductases but lack the NapGH ubiquinol oxidase. *S. oneidensis* MR-1 appear to be atypical because it lacks both *napC* and *nrfBCD*, whose gene products act to transfer electrons from quinol pool to terminal reductases NapA and NrfA. In *E. coli*, NapC and NrfBCD are essential for catalyzing reductions of nitrate to nitrite and the subsequent reduction of nitrite to ammonium, respectively. Our previous results revealed that CymA is likely to be functional replacement of both NapC and NrfBCD in the nitrate and nitrite reduction in *S. oneidensis* MR-1.

The nitrate reduction in W3-18-1 and a closely related strain CN32 were examined, and the transient accumulation of nitrite was also observed in both strains fed with 2mM of nitrate (50 mM of lactate as electron donor), suggesting that the two-step manner of nitrate reduction may be common among *Shewanella* species. A series of mutants in W3-18-1 were generated and tested for the bacterial growth on

nitrate. Deletion of entire operon of *nap1* (*napDAGHB*) or *nap2* (*napDABC*) did not significantly affect the cell growth, but the double mutant with deletion of both *nap* operons, *nap1* and *nap2*, could not grow on nitrate. It is suggested that the two *nap* operons are functionally redundant. In addition, the in-frame deletion mutants of *cymA* and *napC* of W3-18-1 did not show severe growth defectiveness on nitrate, though deletion of *cymA* rendered the loss of reduction of nitrate and nitrite and bacterial growth in MR-1. Furthermore, the *cymA* deletion mutant showed little growth on nitrite in contrast to the *napC* deletion mutant, indicating that CymA was involved in nitrite reduction in both W3-18-1 and MR-1. The *cymA* gene from W3-18-1 could complement the MR-1 *cymA* in-frame deletion mutant in reduction of ferric ions, nitrate, and nitrite when expressed *in trans*. The *napC* gene from W3-18-1 could also complement the defectiveness of MR-1 *cymA* deletion mutant in ferric iron reduction but it failed to complement nitrite reduction. These results support the hypothesis that the complicated periplasmic nitrate reduction systems, NapC-dependent and CymA-dependent, represent a scavenge mechanism for nitrate and nitrite in most *Shewanella* species, which could utilize nitrate and nitrite as both electron acceptors for energy generation and nitrogen nutrient source for cellular metabolisms.

GTL

## The *Shewanella* Knowledgebase

Tatiana Karpinets<sup>1,2</sup> (karpinetsv@ornl.gov),<sup>3</sup> Margaret Romine,<sup>1</sup> Denise Schmoyer,<sup>1,2</sup> Guruprasad Kora,<sup>1</sup> Mike Leuze,<sup>1</sup> Mustafa Syed,<sup>4</sup> Erich Baker,<sup>1</sup> Byung Park,<sup>3</sup> Andrey Gorin,<sup>1,5</sup> Nagiza Samatova,<sup>1</sup> and **Ed Uberbacher<sup>1\*</sup>** (ube@ornl.gov)

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>2</sup>University of Tennessee, Knoxville, Tenn.; <sup>3</sup>Pacific Northwest National Laboratory, Richland, Wash.; <sup>4</sup>Baylor College of Medicine, Waco, Tex.; and <sup>5</sup>North Carolina State University, Raleigh, N.C.

<http://shewanella-knowledgebase.org>

**Project Goals: Provide a framework for data and knowledge integration and sharing for the *Shewanella* Federation.**

The *Shewanella* Knowledgebase is a framework for integrating diverse experimental information on *Shewanella* species with their genomic characteristics. Recent developments in the *Shewanella* Knowledgebase have concentrated in the following areas:

1. Developing Ortholog and Genome Editors for 21 sequenced *Shewanella*
2. Further enhancement of the *Shewanella* Knowledgebase web portal, which is a data and knowledge integration environment that allows investigators to query across the *Shewanella* Federation (SF)

experimental datasets, link to *Shewanella* and other community resources, and visualize the data in a cell systems context.

### 3. Developing a database of SF presentations and reporting materials

*The Ortholog and Genome Editor* is a set of comparative visualization and analytical tools for curation, storage and analysis of genome annotation for all sequenced *Shewanella* (Fig A). This tool is unique in that no other databases or web servers provide the environment to curate orthologous groupings of proteins present across a set of related organisms. The groupings are especially important for comparative analysis of the species and for improving genome annotations of newly sequenced organisms, which are closely related to a better annotated and studied model organism. In case of *Shewanella* spp. the model organism is *Shewanella oneidensis* MR-1, which has significant number of orthologous genes with other species of the genus. By developing the ortholog editor we have achieved the following goals: (i) decreased errors generated by automatic prediction of ortholog, (ii) identified genes, which were missed in the initial annotation, and pseudogenes, (iii) provided environment to store, edit and download the most recent annotations of the sequenced *Shewanella* genomes in different formats, (iv) improved product annotation and made it consistent across all *Shewanella* genomes, (v) make it possible to compare omics data at the gene and pathway level across multiple genomes. In addition to the ortholog table the editor database includes a table of evidence supporting the automatic identification of orthologs. The table includes a set of annotations of the genomes of sequenced *Shewanella* spp using different databases and tools including PROtein K(c) clusters, COG, CDD, Pfam, and SMART. Automatic checks of consistency are implemented for locus tags combined into one orthologous group considering both the length of proteins and their domain structures. The main page of the ortholog editor provides access to the ortholog table, capability for results checking, and the generated alignments (Fig. B). A set of options for table editing, sorting, viewing, searching and downloading are also available. The ortholog editor is linked to individual spp, editors that provide a web interface to store, edit and download in different formats the improved annotations of the sequenced *Shewanella* genomes.

*The Shewanella Knowledgebase web portal* was advanced by collecting experimental data produced by the SF and other researchers with subsequent integration of the data in the knowledgebase environment. A visualization and analytical tool, ShewRegDB, was developed for integration of the regulatory information with metabolic pathway predictions from ShewCyc and with collected experimental information. The Knowledgebase was supplemented by several experimental projects including 3 projects on protein structure and 25 projects involving Affymetrix microarray and two-color microarrays profiling. The regulatory database was enhanced by including 359 binding sites extracted from literature, by independent transcription terminator predictions from TransTermHP, by computational predictions of operons from the DOOR database, MicrobesOnline, ODB,

Shewybase and by manually curated operon annotations. An interface to the regulatory database was developed to search three main data objects: DNA regulator binding sites, RNA regulators and operons. The interface links the regulatory information to information stored in the knowledgebase and provides table overviews and downloads of the search results in EXCEL format. Visualization of the regulatory information is now integrated with the experimental data collected in the knowledgebase and pathway predictions in the updated version of ShewCyc. This information can be displayed through additional tracks in the GBrowser. A Metabolic Pathway Viewer based on Google map technology is under development to track metabolites, overlay experimental data with the metabolic pathways and provide bird eye view of each individual pathway.

*Database of SF members' progress reports* was created providing the following main options: (i) a quick overview of the available reports, presentations, abstracts and posters generated by federation; (ii) downloading all types of reporting materials; (iii) a user friendly interface for uploading a report/presentations/abstracts/poster.

This project is a component of the *Shewanella* Federation and contributes to revealing molecular functions underlying the diverse ecophysiology of this important species.

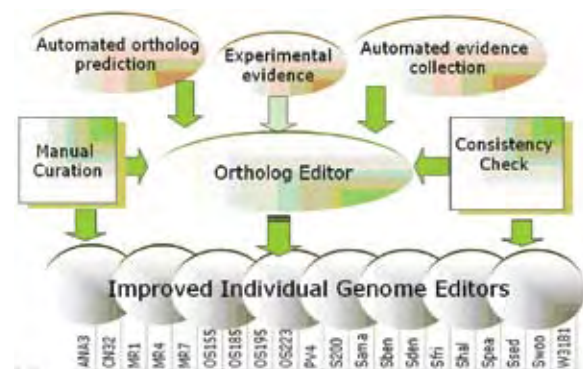


Figure 1: (A)



Figure 2: (B)

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.

## Geobacter Systems Understanding

GTL

### Experimental Genome Annotation of *Geobacter sulfurreducens*

Karsten Zengler,<sup>1</sup> Byung-Kwan Cho<sup>1\*</sup> (bcho@ucsd.edu) Yu Qiu,<sup>1</sup> Christian Barrett,<sup>1</sup> Young-Seoub Park,<sup>1</sup> Mary Lipton,<sup>2</sup> **Derek Lovley**<sup>3</sup> (dlovley@microbio.umass.edu), and Bernhard O. Palsson<sup>1</sup>

<sup>1</sup>Dept. of Bioengineering, University of California San Diego, La Jolla, Calif.; <sup>2</sup>Biosciences Division, Pacific Northwest National Laboratory, Richland, Wash.; and <sup>3</sup>Dept. of Microbiology, University of Massachusetts, Amherst, Mass.

**Project Goals:** 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.

The ability to obtain sequence information and assemble whole genomes, or large parts of it, has been paramount to our understanding of organisms, ranging from bacteria to eukaryotes including humans. Recent advance in sequence technologies have improved the quality, quantity, the time frame in which data can be generated and the cost in ways that have been unimaginable just a few years ago. Generating large amounts of high quality sequence data therefore is not a limiting step in biological sciences any longer. However, when it comes to translating sequences into genes, determine genome structure, and assigning function we rely exclusively on computational methods. Precise annotation of every gene within an organism solely by bioinformatics tools has still not been possible.

Realizing this shortfall there have recently been approaches to experimentally annotate genomes by studying the transcriptome of organisms using high density microarrays and sequencing approaches, respectively. Transcriptome analysis by itself, however, can only answer indirectly what the

functional elements of the genome (transfrags) are and how these elements are organized and regulated—information which is crucial for comprehensive genome annotation. Here we describe the use of high-density tiling microarrays to determine transcripts abundances, novel transcripts, transcription units, and UTR boundaries under different growth conditions. In addition we determined genome-wide binding locations of DNA-binding proteins including RNA polymerase and sigma factors by ChIP-chip. High-density tiling microarrays of overlapping probes can effectively address the transfrags determination since they are constructed without any *a priori* knowledge of the possible transfrag content of a genomic sequence. However, for an experiment-based annotation it is still necessary to pinpoint the accurate structures (i.e., the boundaries of UTRs and transcription start site) of transcription units (TU). In order to map the transcription start sites of the bacterial genomes, we developed a modified mRNA-seq method using RNA ligation followed by Solexa sequencing. Whether novel transcripts were encoding for proteins or maybe non-coding RNAs was addressed by comprehensive proteomic analysis (>85% coverage of formerly predicted coding proteins). Integration of all these genome wide data sets allowed for a comprehensive experimental genome annotation including several new transcripts and non-coding RNAs.

GTL

### Novel Approaches for Genome-Scale Spatial Analysis of Gene Transcription in Biofilms: Elucidation of Differences in Metabolism Throughout *Geobacter sulfurreducens* Biofilms Producing High Current Densities

Ashley E. Franks\* (aefranks@microbio.umass.edu), Kelly P. Nevin, Kengo Inoue, Richard H. Glaven, and **Derek R. Lovley**

Dept. of Microbiology, University of Massachusetts, Amherst, Mass.

<http://www.geobacter.org>

**Project Goals:** 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.**

Microbial fuel cells (MFCs) show promise as a strategy for converting wastes and biomass directly to electrical current, providing a new form of bioenergy. Studies in a number of laboratories have demonstrated that, in the most highly effective MFCs *Geobacter sulfurreducens*, or closely related organisms, outcompete the other diverse constituents of complex microbial communities to preferentially colonize the current-harvesting anodes. Therefore, further information on the growth and metabolism of *G. sulfurreducens* on anodes is expected to lead to the design of superior MFCs. Our previous studies have demonstrated that *G. sulfurreducens* produces thick (>50  $\mu\text{m}$ ), electrically conductive biofilms on MFC anodes in systems designed for high current densities. It was hypothesized that environmental conditions might vary significantly throughout the biofilm and, in some instances, might result in suboptimal physiological states that could negatively impact on current production.

Therefore, a MFC that permits real time imaging of the anode biofilm with confocal scanning laser microscopy was developed. *G. sulfurreducens* produced current in this new system at rates comparable to that previously reported with other MFC designs. Cells engineered to produce the fluorescent protein mcherry to facilitate real-time imaging produced current comparable to wild-type cells. Metabolic staining of a current-producing biofilm demonstrated metabolic activity across the entire thickness of the electrically conductive biofilm. Introducing C-SNARF-4, a pH-sensitive fluorophore, into the anode chamber revealed strong pH gradients within the anode biofilms. The pH decreased with increased proximity to the anode surface and from the exterior to the interior of biofilm pillars. Near the anode surface pH levels were as low as 6.1 compared to ca. 7 in the external medium. Various controls demonstrated that the proton accumulation was associated with current production. These results demonstrated that it is feasible to non-destructively monitor the activity of anode biofilms in real time and suggest that the accumulation of protons that are released from organic matter oxidation within anode biofilms can limit current production.

In order to investigate levels of gene transcription throughout the biofilm, reporter plasmids were created by cloning short half-life fluorescent proteins into the plasmid pRG5. Promoters of genes of interest were inserted upstream of the fluorescent proteins, controlling their expression. This makes it possible to track gene expression at different layers of the biofilm in real time with confocal microscopy. The short half-life of the fluorescent proteins permits monitoring of repression of gene expression as well as induction. Initial studies are focusing on expression of the gene for the enzyme citrate synthase because our previous studies with chemostat-grown *G. sulfurreducens* have demonstrated that expression levels of this gene are related to rates of metabolism. Other reporters will include those for the gene, *pilA*, which encodes the structural protein for the electri-

cally conductive, microbial nanowires, and for cytochromes that contribute to the electrical conductivity of the biofilm and/or serve as electrical contacts between the cells and the anode.

Another strategy that we developed for investigating gene expression throughout the biofilm involved a novel biofilm sectioning technique coupled with whole-genome microarray analysis of transcript abundance. Biofilms were treated with RNase protect, hardened with resin, and sliced horizontally in 100 nm slices with a diamond knife. Slices were pooled into inner and outer portions, the RNA extracted and transcript abundance analyzed with microarrays. There were 146 genes which were differentially expressed (2-fold cutoff;  $p < 0.05$ ) between the inner and outer portions of the biofilm. Only 1 gene, encoding a putative ABC transporter, was up-regulated more than 2-fold in the outer portion of the biofilm. Of the genes down regulated in the outer portion of the biofilm, many had to do with cell metabolism and growth, consistent with the concept that cells at a distance from the anode have lower rates of metabolism.

Recent studies on the outer-surface *c*-type cytochrome, OmcZ, illustrate how such gene expression studies can lead to enhanced understanding of the functioning of *G. sulfurreducens* in anode biofilms. Microarray analysis indicated that *omcZ* was much more highly expressed in current-producing biofilms than in biofilms growing on the same electrode material, but using fumarate as the electron acceptor. Gene deletion and complementation studies demonstrated that *omcZ* was essential for high-density current production. Electrochemical analysis demonstrated that the biofilms of the *omcZ*-deficient strain were highly resistive. OmcZ was purified and characterized. Preliminary localization studies with gold-labeled antibodies demonstrated that OmcZ was extracellular and dispersed throughout *G. sulfurreducens* biofilms. These results suggest that OmcZ enables long-range electron transfer through *G. sulfurreducens* anode biofilms by providing a mechanism for electron transfer through the biofilm matrix. If so, then increasing expression of OmcZ may improve biofilm conductivity and hence MFC current densities.

These studies represent the first analysis of the gene expression and physiological status of high current density biofilms. The microtoming studies represent the first genome-scale analysis of gene transcript abundance within individual depths of a biofilm of any type. The results from these studies are providing new insights into the function of MFC anode biofilms that are expected to lead to better strategies for optimizing the power output of MFCs and thus broaden their applications.

GTL

## Genome-Wide Mapping of Transcriptional Start Sites of *Geobacter sulfurreducens* using High-Throughput Sequence Methodologies

Katy Juárez<sup>1\*</sup> (katy@ibt.unam.mx), Alfredo Mendoza,<sup>1</sup> Leticia Vega-Alvarado,<sup>2</sup> Blanca Taboada,<sup>2</sup> Verónica Jiménez,<sup>3</sup> Leticia Olvera,<sup>1</sup> Maricela Olvera,<sup>1</sup> Ana L. Tirado,<sup>1</sup> Julio Collado-Vides,<sup>3</sup> Enrique Morett,<sup>1</sup> and **Derek R. Lovley**<sup>4</sup> (dlovley@microbio.umass.edu)

<sup>1</sup>Instituto de Biotecnología, <sup>2</sup>Centro de Ciencias Aplicadas y Desarrollo Tecnológico and <sup>3</sup>Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico; and <sup>4</sup>Dept. of Microbiology, University of Massachusetts, Amherst, Mass.

**Project Goals:** The goal of this project is to provide computational tools to predicatively 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.

U(VI) bioremediation strategies require an extensive understanding of metal-reducing microorganisms, the environmental parameters controlling their metabolism and the regulation of their gene expression. Molecular analyses have clearly indicated that *Geobacteraceae* are the predominant bacteria of the microbial community during in situ bioremediation of uranium-contaminated environments and when dissimilatory metal reduction is stimulated in the subsurface. Although significant progress has been made in the knowledge of the electron transfer mechanisms in *Geobacter sulfurreducens*, little is known about the regulatory cues involved in controlling gene expression of the participants in this complex process.

Gene expression starts by the specific and normally tightly regulated transcription initiation process. The experimental determination of the genome-wide transcription start sites (TSS), will contribute, not only to improve the knowledge about the promoter and operon structure in this bacterium, but also to increase the predictive capacity for those cases

not experimentally determined. Having precise information about the majority of the TSSs in *G. sulfurreducens* will certainly help to achieve a detailed understanding of the global regulatory circuits that control gene expression in this organism.

In order to obtain a global picture of the active promoters in *G. sulfurreducens* growing in different environmental conditions, we have developed two approaches to map thousands of transcriptional start sites: one is directed mapping using a modification of the 5'RACE protocol, and the other one is a global mapping of TSS using 454 pyrosequencing technology. Using these strategies in *G. sulfurreducens*, we have been able to map hundreds of new TSS and confirm some TSS previously reported, indicating that our methodology is accurate and robust. We also developed a web-based tool to analyze the results of the pyrosequencing reactions and the information obtained is available at <http://geobactertss.ccg.unam.mx> and it has allowed us to analyze very efficiently the individual sequences over the *G. sulfurreducens* genome.

The results of the first cDNA library show a distribution throughout the genome and in correlation with the gene expression levels. We are evaluating the possible occurrence of intragenic promoters and new coding small ORFs. Interestingly, in this first collection we identified additional TSS for some genes previously studied. This work adds substantial new data to our ongoing large-scale effort to experimentally determine TSS, promoters and regulatory elements in *G. sulfurreducens*. We have found that it is not uncommon that genes are expressed from multiple promoters, which certainly enriches the regulatory options to control gene expression. Our results provide new information that certainly will facilitate the understanding from a global perspective the complex and intricate regulatory network that operates in *G. sulfurreducens*.

GTL

## Genome Resequencing Reveals that Current-Harvesting Electrodes Select for Rare Variant of *Geobacter sulfurreducens* Capable of Enhanced Current Production

Anna Klimes<sup>1\*</sup> (aklimes@microbio.umass.edu), Kelly P. Nevin,<sup>1</sup> Hana Yi,<sup>1</sup> Jessica E. Butler,<sup>1</sup> Trevor L. Woodard,<sup>1</sup> Christian L. Barrett,<sup>2</sup> Karsten Zengler,<sup>2</sup> Bernhard Palsson,<sup>2</sup> Barbara Methe,<sup>3</sup> Yuan Gao,<sup>4</sup> Bin Xie,<sup>4</sup> Chen Liang,<sup>5</sup> and **Derek R. Lovley**<sup>1</sup>

<sup>1</sup>Dept. of Microbiology, University of Massachusetts, Amherst, Mass.; <sup>2</sup>Dept. of Bioengineering, University of California San Diego, La Jolla, Calif.; <sup>3</sup>J. Craig Venter Institute, Rockville, Md.; <sup>4</sup>Virginia Commonwealth University, Richmond, Va.; and <sup>5</sup>Beijing University of Aeronautics and Astronautics, Beijing, China

**Project Goals:** 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.**

Electricity is not commonly considered as a form of bioenergy, but microbial fuel cells (MFCs) can effectively convert a range of organic wastes and biomass directly to current. At present, the application of MFCs is limited by low power outputs. One likely reason for the low power is that electron transfer to electrodes is probably only a fortuitous process, related to extracellular electron transfer to natural extracellular electron acceptors such as Fe(III) oxides. As far as is known, there has been no previous selective pressure on microorganisms to produce current. Thus, this process is probably far from optimized.

*Geobacter sulfurreducens* is capable of producing current densities as high, or higher, than any other known pure or mixed culture. In an attempt to develop a strain of *G. sulfurreducens* with improved current production capability we repeatedly transferred cells grown on the current-harvesting anodes of microbial fuel cells, or selected for growth at low anode potentials, with the assumption that this selective pressure would result in the accumulation of beneficial mutations that would enhance the capacity for electron transfer to anodes. Strains were obtained that produced up to 30-fold more power than the starting culture.

Resequencing the genomes of twelve of these electrode-selected strains with Illumina sequencing technology revealed that the sequences of the 16S rRNA operons were identical to the starting culture, indicating the electrode-selected strains were *G. sulfurreducens*. However, each strain had over 18,000 nucleotide mutations compared with the genome sequence obtained from the starting culture. Furthermore, the electrode-selected strains appeared to be missing ca. 5% of the sequence found in the starting culture. Similar results were obtained with Nimblegen re-sequencing chip technology. About a third of the genes in the genome contained at least one nucleotide mutation. The mutations were found throughout the length of the genome, with some genes containing over 100 mutations. There was a heavy bias for the mutations to be in the third position of the codon, which resulted in the majority of the mutations (ca. 75%) causing synonymous changes, meaning no change in the protein sequences. Furthermore, the electrode-selected

strains contained ca. 100 kb of DNA sequence that was not detected in sequencing of the starting culture.

This degree of genome change is much greater than can be attributed to random mutation and genome rearrangements during the course of selection on the electrodes. Rather, the results suggest that the starting culture of *G. sulfurreducens* contains at least two strains: the strain that predominates under all previously described culturing conditions (referred to here as strain DL1) and at least one variant that is specifically selected during growth on electrodes. The fact that the variant was not detected in the original culture even after 80-fold coverage of the genome sequence with Illumina sequencing indicates that the electrode-specific variant is in very low abundance under normal culturing conditions. The most heavily mutated genes, common to all sequenced strains of the variant, are predicted to have roles in energy metabolism, cell envelope synthesis, and transport. Several TCA cycle proteins and several clusters of cytochromes were heavily mutated, which may affect electron transport pathways. In addition, there were nonsynonymous mutations in proteins required for DNA synthesis and replication, and for DNA mismatch and strand break repair, suggesting that the electrode-selected variants might have a higher mutation rate than the DL1 strain.

PCR primers designed to amplify four of the genes found in the variants, but not in the DL1 strain, amplified the correct PCR product from the original *G. sulfurreducens* culture. However, a nested PCR protocol was required, suggesting that the variant was in very low abundance in the culture. This low abundance was verified with quantitative PCR and is consistent with results of Illumina sequencing of the starting culture.

In addition to their enhanced capability for current production, the electrode-selected variants had other unique features that distinguished them from the DL1 strain. For example, the electrode-selected variants were highly motile, with long flagella, whereas DL1 is non-motile and does not produce flagella. Compared with DL1, the electrode-selected strains had a greater abundance of pilin-like filaments that look like the electrically conductive pili, known as microbial nanowires. Furthermore, analysis of one of the electrode-selected strains demonstrated that its biofilms are much more conductive than DL1 biofilms. Electrochemical analysis of the anode biofilms of this strain demonstrated that it had a lower mid-point potential than DL1 for electron transfer to electrodes, consistent with its ability to produce much higher current on electrodes poised at low potentials. This strain also had a greater capacity to adhere to glass or graphite. This was associated with major changes in lipopolysaccharide structure and exopolysaccharide content. Further comparative genome-scale studies of these strains are expected to yield additional insights into the features that contribute to enhanced extracellular electron transfer.

These results have obvious implications for enhancing the power output of microbial fuel cells. Furthermore, these findings demonstrate that cultures that appear to be “pure”, even with the most advanced sequencing techniques, may

in fact contain variants, which have substantial differences in genomic content and physiological capabilities, that can only be detected by forcing the culture through an extreme bottleneck that strongly favors the growth of the rare variant. This phenomenon may be an important consideration for the interpretation of other adaptation studies.

—  
GTL

## Bioinformatic Analysis of Gene Regulation in *Geobacter sulfurreducens*: an Integration of Transcriptome and Sequence Information, Molecular Evolutionary Studies, and Database Management

Julia Krushkal<sup>1\*</sup> (jkrushka@utmem.edu), Yanhua Qu,<sup>1</sup> Peter Brown,<sup>2</sup> Sreedhar Sontineni,<sup>1</sup> Toshiyuki Ueki,<sup>2</sup> Katy Juarez,<sup>3</sup> Ching Leang,<sup>2</sup> Enrique Merino,<sup>3</sup> Jeanette Peebles,<sup>1</sup> Jose F. Barbe,<sup>1</sup> Ronald M. Adkins,<sup>1</sup> and **Derek R. Lovley** (dlovley@microbio.umass.edu)

<sup>1</sup>University of Tennessee Health Science Center, Memphis, Tenn.; <sup>2</sup>University of Massachusetts, Amherst, Mass.; and <sup>3</sup>Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, México

**Project Goals: 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.**

Understanding regulation of gene expression in the delta-proteobacterial family *Geobacteraceae* is critical for our ability to gain insight into the cellular processes which allow these bacteria to participate in environmental bioremediation and energy production. To date, a significant amount of experimental data has been accumulated, providing information about the genome structure of these organisms and describing the changes in their expression at the transcriptome and the proteome level. We have been employing bioinformatic approaches to identify regulatory interactions by synthesizing available knowledge on transcriptional regulators and

their predicted and/or experimentally identified target genome sites gained from these experimental studies.

To address the need for a systematic, comprehensive resource integrating available regulatory information, we have developed an online database, GSEL (*Geobacter* Sequence Elements), which compiles regulatory information for *Geobacter sulfurreducens*, an intensively studied model representative of *Geobacteraceae* whose regulation is currently best understood. We have recently completed the development of a new, significantly expanded and updated, relational version 2 of the GSEL database and its accompanying online query system, which compiles manually curated information on operon organization and transcription regulatory elements in the genome of *G. sulfurreducens*. GSEL v. 2 incorporates a new online graphical browser, and it provides significantly expanded search capabilities allowing users to query the database to identify and view graphically the operon structure and regulatory sequence elements in a genome region of interest. Users can search the database by providing genome coordinates, operon ID, or gene ID. Users can also identify transcriptional regulatory sites recognized by a specific transcriptional regulator, those generated in a specific microarray experiment, or those predicted using a specific search method. This new version of the GSEL database includes updated operon predictions along with new information on predicted and/or experimentally validated genome regulatory sites, including promoters, transcriptional factor binding sites, transcriptional attenuators, ribosome-binding sites, and terminator sequences. New features in GSEL v. 2 also include links to information from microarray experiments stored in public gene expression databases (ArrayExpress and Gene Expression Omnibus), and to original publications in PubMed or in other bibliography resources describing how particular regulatory interactions were identified. The development of the relational GSEL database version 2 and its online query system was performed using the tools of LAMP (Linux, Apache, MySQL, and PHP) web development, along with components developed using Java, XML, and JavaScript. The GSEL database version 2, its accompanying online search system and a graphic genome browser provide a unique and comprehensive tool cataloguing information about transcription regulation in *G. sulfurreducens*, which aids in the investigation of mechanisms that regulate its ability to generate electric power, bioremediate environmental waste, and adapt to environmental changes.

In addition to continuing development of database resources, we are also continuing our studies to understand regulatory processes which affect the ability of *Geobacteraceae* to participate in energy production. In our earlier studies, we investigated in detail the target genes and promoters regulated by RpoN, an alternative RNA polymerase sigma factor, which regulates a variety of important cellular processes in *G. sulfurreducens*. We identified multiple binding sites for RpoN upstream of genes encoding components or regulators of flagellar biosynthesis, ion transport, nitrogen metabolism, signal transduction, multiple *c*-type cytochromes, members of ABC-type branched-chain transporter system, and other important processes and cell systems.

The most highly conserved RpoN-regulated promoter was located upstream of an operon containing a gene for flagellar transcription factor FliA ( $\sigma^{28}$ ) and other genes for flagellar biosynthesis. Combined with the absence of the master regulator FlhCD in *Geobacter sulfurreducens*, these results strongly indicate the importance of the *Geobacter sulfurreducens* RpoN sigma factor in flagellation.

Our current studies are focusing on an investigation of the roles of several transcription regulatory systems that are likely to be involved in RpoN-dependent regulatory pathways. In particular, we have investigated the target regulatory sites for an enhancer binding protein, PilR, which participates in RpoN-dependent transcriptional regulation of the *pilA* gene encoding structural pilin. We have predicted multiple PilR-regulated sites upstream of multiple operons related to biosynthesis, assembly, and function of pili and flagella, type II secretory pathways, and cell wall biogenesis. A number of these sites have been found to co-occur with RpoN-regulated promoters, providing a further insight into the regulatory role of RpoN in members of the PilR regulon. We have also been investigating sequence changes and molecular classification of the TetR family of transcriptional regulators in *Geobacteraceae*. In *Geobacter sulfurreducens*, we identified RpoN-regulated promoters upstream of several operons containing genes encoding TetR family regulators. The genome of *G. sulfurreducens* contains nine genes encoding TetR family members. Some of these genes are located immediately upstream of operons encoding functionally important *c*-type cytochromes, e.g., *omcB* and *omcC*, which have been suggested to be actively involved in iron reduction. In order to better understand the role played by the TetR family of transcriptional regulators in the ability of *Geobacteraceae* to participate in electron transfer, we have undertaken a systematic molecular evolutionary study by inferring phylogenetic relationships of the *Geobacteraceae* TetR proteins to their homologs in other species of bacteria. This has allowed us to identify conserved and variable domains in these *Geobacter* proteins, which may be important for the diversity of their functional roles, and to classify them into subgroups based on their sequence similarities. We are currently focusing on the classification of molecular changes in these proteins that distinguish them from their close homologs. Combined with our earlier bioinformatic research on RpoS- and RpoD-dependent regulation in *Geobacter sulfurreducens*, these studies are allowing us to obtain a systematic, detailed view of specific regulatory components affecting transcriptional regulation of electron transfer pathways in *Geobacteraceae*.

## Dynamic Genome-Scale Modeling of *Geobacter* Species in Subsurface Environments

R. Mahadevan,<sup>2\*</sup> Jun Sun,<sup>1</sup> Kai Zhuang,<sup>2</sup> Yilin Fang,<sup>3</sup> Tim Scheibe,<sup>3</sup> Hanno Richter,<sup>4</sup> Carla Risso,<sup>4</sup> Paula Mouser,<sup>4</sup> Ashley Franks,<sup>4</sup> and Derek Lovley<sup>4</sup>

<sup>1</sup>Genomatica Inc., San Diego, Calif.; <sup>2</sup>University of Toronto, Toronto, Ontario; <sup>3</sup>Pacific Northwest National Laboratory, Richland, Wash.; and <sup>4</sup>University of Massachusetts, Amherst, Mass.

**Project Goals:** 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.

Previous studies of the *in situ* bioremediation of uranium-contaminated groundwater have demonstrated that *Geobacter* species are the dominant members of the groundwater community during active bioremediation and the primary organisms catalyzing U(VI) reduction. The addition of acetate to the subsurface to stimulate U(VI) reduction has been shown to promote the growth of *Geobacter* species. Fe(III) serves as the primary electron acceptor and U(VI) is simultaneously reduced. However, the dynamics of growth and activity of *Geobacter* species and their interactions with other environmental factors such as groundwater flow and geochemistry are not well understood. Furthermore, the *Geobacter* species cooperate and compete with other subsurface microorganisms and these interactions are also expected to influence the effectiveness of bioremediation. Hence, the objectives of this program were to develop comprehensive metabolic models of *Geobacter* species and other environmentally relevant microorganisms and to integrate these models with descriptions of other physico-chemical phenomena in order to better understand and optimize bioremediation strategies.

For example, recent molecular ecology studies have demonstrated that *Geobacter* species must compete with acetate-oxidizing, Fe(III)-reducing *Rhodospirillum rubrum* species in the subsurface. Prior to the addition of acetate to stimulate U(VI)



reduction, *Rhodoferrax* competes well with *Geobacter* species at the uranium-contaminated field study site in Rifle, CO. In fact, in zones where substantial quantities of ammonium are available, *Rhodoferrax* species are more prevalent than *Geobacter* species. In order to expand our limited knowledge about Fe(III)-reducing *Rhodoferrax* species, *R. ferrireducens* was studied with a multi-faceted approach that included the description and analysis of the genome annotation, investigation of new metabolic capabilities, and the development of a constraint-based *in silico* metabolic model. The model provided new insights into the stoichiometry of the electron transport chain and the efficiency of substrate utilization under different conditions.

The newly developed *R. ferrireducens* genome-scale model was coupled with our previously developed *G. sulfurreducens* model in order to understand the competition between these organisms before and during *in situ* bioremediation of uranium-contaminated subsurface environments. This also required characterization of the acetate transport systems and their uptake kinetics. The simulation of the competition between these organisms suggested that the competition between them is modulated by two factors: the ability of *G. sulfurreducens* to fix nitrogen under ammonium limitation, and a rate vs. yield trade-off between these two organisms. Prior to acetate amendment, if ammonium is limited, *G. sulfurreducens* dominates due to its ability to fix nitrogen. However, if the system contains abundant ammonium, *R. ferrireducens*, which has higher biomass yields, is favoured because the acetate flux is very low. During acetate amendment, the high acetate flux strongly favours *G. sulfurreducens*, a rate-strategist. The model also predicts high respiration rates by *G. sulfurreducens* during nitrogen fixation at the expense of biomass yield, leading to an increase in U(VI) reduction under low ammonium conditions. The results of the simulation agreed well with the subsurface community composition determined from molecular analysis of subsurface samples before and during *in situ* bioremediation at the Rifle site.

In an attempt to better understand and predict how *Geobacter* species might grow and metabolize in the subsurface under diverse conditions that might be imposed in different bioremediation strategies the genome-scale, constraint-based model of the metabolism of *G. sulfurreducens* was coupled with the reactive transport model HYDRO-GEOCHEM. The initial modeling was simplified by only considering the influence of three growth factors: acetate, the electron donor added to stimulate U(VI) reduction; Fe(III), the electron acceptor primarily supporting growth of *Geobacter*; and ammonium, a key nutrient. The constraint-based model predicted that growth yields of *Geobacter* varied significantly based on the availability of these three growth factors and that there are minimum thresholds of acetate and Fe(III) below which growth and activity are not possible. This contrasts with typical, empirical microbial models which assume fixed growth yields and the possibility for complete metabolism of the substrates. The coupled genome-scale and reactive transport model predicted acetate concentrations and U(VI) reduction rates in a field trial of *in situ* uranium bioremediation that were comparable to the

predictions of a calibrated conventional model, but without the need for empirical calibration required for conventional modeling, other than specifying the initial biomass of *Geobacter*. These results suggest that coupling genome-scale metabolic models with reactive transport models may be a good approach to developing models that can be truly predictive, without empirical calibration, for evaluating the probable response of subsurface microorganisms to possible bioremediation approaches prior to implementation.

One of the limitations of the acetate-amendment strategy is that over time, Fe(III) oxides are depleted near the site of acetate injection and acetate-oxidizing sulfate reducers, which are ineffective in U(VI) reduction, become predominant, consume the acetate, and the effectiveness of the bioremediation deteriorates. It has been proposed that this problem might be circumvented by adding lactate instead of acetate with the hope that lactate would stimulate the growth of the lactate-oxidizing sulfate-reducing *Desulfovibrio* species which are highly effective U(VI) reducers. Furthermore, *Desulfovibrio* species only incompletely oxidize lactate to acetate, and thus acetate generated from the metabolism of lactate could potentially support the growth of *Geobacter* species, but the growth of acetate-oxidizing sulfate reducers would be prevented because *Desulfovibrio* would have consumed the sulfate. Before incurring the cost and time of implementing this new bioremediation strategy it is desirable to evaluate its likely outcome *in silico*. Therefore, we developed a computational model of *Geobacter*, *Desulfovibrio*, and the acetate-oxidizing sulfate-reducing bacterial community by dynamically integrating the genome-scale metabolic model of *G. sulfurreducens*, a central metabolic model of *D. vulgaris*, and the kinetic model of a representative acetate-oxidizing sulfate reducer using the dynamic multi-species metabolic modeling framework. Simulations of batch, continuous, and fed-batch lactate injection indicated that the syntrophic community based on acetate exchange between *D. vulgaris* and *G. sulfurreducens* can lead to prolonged Fe(III) reduction and can potentially accelerate U(VI) reduction. The next step will be to couple this model with the reactive transport model to design optimal feed strategies for this bioremediation approach.

These studies demonstrate that genome-scale modeling can aid in the interpretation and prediction of microbe-microbe and microbe-geochemical interactions in the subsurface. In addition to enhancing the understanding of basic subsurface microbiology, this approach has clear practical application for bioremediation optimization.

## The Application of Metagenomic and Metatranscriptomic Approaches to the Study of Microbial Communities in a Uranium-Contaminated Subsurface Environment

Barbara Methé<sup>1\*</sup> (bmethe@jcvi.org), Robert Deboy,<sup>1</sup> Qinghu Ren,<sup>1</sup> Michael Geimer,<sup>1</sup> Jamison McCarrison,<sup>1</sup> Dawn Holmes,<sup>2</sup> and Derek Lovley<sup>2</sup>

<sup>1</sup>The J. Craig Venter Institute, Rockville, Md.; and

<sup>2</sup>University of Massachusetts, Amherst, Mass.

**Project Goals:** 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.

The Old Rifle site located in Rifle, Colorado represents a former uranium ore-processing facility which is currently being managed as a part of the Uranium Mill Tailings Remedial Action (UMTRA) program of the U.S. Department of Energy. Over time, continuous leaching of radionuclides from processed mill tailings present at this location resulted in contamination of both the groundwater and sediment of the surrounding aquifer. Currently, in situ bioremediation pilot studies, specifically the biostimulation of members of the *Geobacteraceae* through the addition of acetate to groundwater, is being investigated at this site as a possible corrective strategy. The study presented here is meant to complement these wider ongoing investigations through the application of next generation sequencing and metagenomic analyses in an effort to reveal genetic information from subsurface microorganisms without the need for prior cultivation.

A metagenomic framework for use in this project has been established as follows. Genomic DNA from the 2007 sampling of the Rifle D05 site groundwater was used to prepare a 3–4 kb insert paired-end library for sequencing using 454 FLX pyrosequencing. One run of 454 sequencing has been completed yielding ~35% paired-end reads and 65% fragments (single reads) which generated 127,849,360 total

bases of an average read length of 245bp and an average quality value of 33.

Initial analyses of this data have focused on the evaluation of a variety of computational methods to investigate taxonomic and functional classifications of both the unassembled and assembled reads. Assemblies were completed using the Newbler Assembler from 454 and a modified version of the Celera Assembler (CA). A greater N50 contig and scaffold size, respectively (N50 contig= 2419 vs. 1339 and N50 scaffold= 19098 vs. 15586) was achieved with CA. In addition, results of whole genome alignments using Nucmer demonstrated greater coverage of reference genomes from the *Geobacteraceae* with the CA versus the Newbler data set. Therefore, the CA assembly has been used for additional analyses of assembled data.

Taxonomic classification of individual reads and contigs was completed using the Metagenome Analyzer (MEGAN) program. Results from this analysis identified ~57% of all reads assigned at the Family level were classified as members of the *Geobacteraceae* (*Geobacteraceae* + *Pelobacteraceae* based on NCBI taxonomy). From an examination of assembled contigs, approximately 55% were assigned to the genus *Geobacter* with members of the Subsurface Clade 1 including *G. sp. M21*, *G. bemidjiensis* and *G. uraniumreducens* representing the predominate species classifications.

Functional classifications of the individual 454 reads were performed using the MG-RAST pipeline. In this analysis a total of 499,050 reads representing ~121 Mb were examined of which 195,949 sequences (39%) could be matched to proteins in SEED subsystems using an e-value cut-off of  $1e^{-5}$ . The top five functional categories determined were: Carbohydrate Metabolism (11.6%), Clustering-based Subsystems (10.6%), Virulence (9.4%), Amino Acids and Derivatives (7.5%) and Protein Metabolism (7.1%). Carbohydrate Metabolism was dominated by reads placed in the subcategory of central intermediary metabolism (~24% of the total for the category) but reads related to functions such as biofilm formation and, methanogenesis and one-carbon metabolism, were also found. The category of Virulence was dominated by reads classified in the subcategories of resistance to antibiotics and toxic compounds (36%), and prophage and transposons (29%). However, classification of reads in subcategories such as Type IV pilin as well as Type III and Type VI secretion systems which are often found in pathogenic organisms, but more recently have been identified in non-pathogenic bacteria, were also noted.

From the assembled data, the program, Metagene, was used to predict ORFs from the contigs generated by CA. A total of 8105 ORFs consisting of 4274 complete and 3831 partial ORFs were predicted. Analysis of annotation of these ORFs has suggested that even though the majority of assembled genomic data is related to members of the *Geobacteraceae*, many ORFs (at least one-third) identified are not closely related to sequenced organisms in culture. These results suggest that even among the *Geobacteraceae*, a greater diversity of organisms exists at this site than has currently been captured in pure culture and sequenced.

In addition to examining the D05 microbial community at the level of the genome, we have also begun experimentation designed to elucidate community gene expression through investigation of the metatranscriptome. We have used a capture oligonucleotide approach to enrich mRNA through the removal of 16S or 23S rRNA molecules. The enriched mRNA is then used as template in cDNA reactions which can subsequently be sequenced via 454 pyrosequencing. Currently we have produced cDNA from enriched mRNA and have tested the efficacy of the cDNA through a series of RT-PCR-based experiments targeting a set of housekeeping genes that are generally constitutively expressed. Included in this series of validations has been the design of primer sequences to these housekeeping targets based on the metagenomic sequence. From D05 cDNA we have used these primers to amplify a region of the target gene and validated these results by sequencing and analysis of the gene product.

Additional investigations are now underway to further examine the D05 metagenome. These include the use of: 1) an iterative protein clustering approach to further identify and characterize protein families from individual reads and predicted ORFs, and 2) the application of a web based tool developed at the JCVI, Advanced Reference Viewer, designed to facilitate comparative analysis of metagenomic datasets in the context of a reference genome and 3) comparison of the metatranscriptome (gene expression) to the metagenome. Overall, the efforts presented here are central to several goals of this project. Since the samples examined were obtained from bulk community DNA and RNA as opposed to individual populations, the resulting information may serve to generate new insights and hypotheses concerning metabolic processes and interactions between community members. Further, these results may also serve as a framework for predictive modeling of processes relevant to bioremediation.

**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.**

The Fe(III)-reducing bacterium *Rhodoferrax ferrireducens* is a metabolically versatile organism known for its ability to convert sugars to electricity and its important role in dissimilatory metal reduction in some subsurface environments. In order to expand our limited knowledge about *R. ferrireducens*, a multi-approach study was conducted, including the description and analysis of the genome annotation, the development of a constraint-based genome-scale metabolic model, and physiological studies.

The *R. ferrireducens* genome consists of a circular chromosome of 4,712,337 base pairs and a plasmid with 257,447 bp. A total of 4,770 protein-encoding open reading frames (ORFs) have been predicted in the chromosome, of which 2,797 have been assigned a putative function. Analysis of the top BLASTP hits of *R. ferrireducens* ORFs against a database of complete genomes established the following taxonomic breakdown: *Betaproteobacteria* (3,253 genes), *Gammaproteobacteria* (258), *Alphaproteobacteria* (185), *Delta/Epsilonproteobacteria* (121). Of completed genomes, those representing *Polaromonas* sp. are the most closely related to *R. ferrireducens*, followed by *Acidovorax* sp., *Verminephrobacter eiseniae*, *Methylobium petroleiphylum* and *Ralstonia* sp. There are 169 *R. ferrireducens* genes whose best match is to another *R. ferrireducens* gene suggesting these might be lineage-specific duplications. Conversely, about 14% of the *R. ferrireducens* genes are hypothetical ORFs.

The genome annotation of *R. ferrireducens* was used for the construction of the metabolic network. The annotated genes of the *R. ferrireducens* genome, as well as genes from several high-quality genome-scale metabolic models, were utilized to generate a draft network as a starting point for model reconstruction. The draft model successfully captured significant portions of central metabolism, as well as biosynthetic pathways for amino acids, nucleotides, and lipids. Among the base models used, *E. coli* was the phylogenetically closest to *R. ferrireducens* and provided about half of all reactions in the draft model. Of the 4770 genes in the *R. ferrireducens* genome, 744 genes were included in the reconstructed genome-scale network. The *R. ferrireducens* metabolic model contains 762 reactions and 790 metabolites including 69 extracellular metabolites. The metabolic capabilities of the *R. ferrireducens* network were calculated using flux balance analysis and linear optimization. Biomass synthesis was

—  
GTL

## New Insights into *Rhodoferrax ferrireducens* Through Genome Annotation and Genome-Based *in silico* Metabolic Modeling

Carla Risso<sup>1\*</sup> (crisso@microbio.umass.edu), Jun Sun,<sup>2</sup> Kai Zhuang,<sup>3</sup> Radhakrishnan Mahadevan,<sup>3</sup> Olivia Bui,<sup>2</sup> Robert DeBoy,<sup>4</sup> Wael Ismael,<sup>1</sup> Christophe H. Schilling,<sup>2</sup> Barbara A. Methé,<sup>4</sup> and Derek Lovley<sup>1</sup> (dlovley@microbio.umass.edu)

<sup>1</sup>Dept. of Microbiology, University of Massachusetts, Amherst, Mass.; <sup>2</sup>Genomatica Inc., San Diego, Calif.; <sup>3</sup>University of Toronto, Canada; and <sup>4</sup>J. Craig Venter Institute, Rockville, Md.

**Project Goals: 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**

selected as the objective function to be maximized in growth simulations, and ATP consumption was selected as the objective function to be maximized in energy requirement simulations.

The *in silico* model was used to estimate energy parameters. An optimization algorithm iterated the  $H^+/2e^-$  ratio of NADH dehydrogenase from 1 to 4, the  $H^+/2e^-$  ratio of cytochrome reductase from 1 to 4, and non-growth associated maintenance (nGAM) from 0 to 2.5 mmol ATP/gDW/hr to estimate the values that would result in the most consistent growth yields between those experimentally determined versus those predicted by the model. This process identified optimal energy parameters of an  $H^+/2e^-$  ratio of 2 for both NADH dehydrogenase and cytochrome reductase, and an nGAM of 0.45 mmol ATP/gDW/hr. This set of energy parameters was applied to the model and validated by the comparison between *in silico* predictions and experimentally determined yields.

Several previously unknown metabolic capabilities of *R. ferrireducens* were discovered. These included the ability to utilize carbohydrates other than glucose to support growth, including fructose and mannose. The catabolism of certain disaccharides was also determined. Notably, cellobiose degradation poses biotechnological interest because of the potential of turning common cellulosic waste products into energy. It was found that *R. ferrireducens* is capable of fumarate dismutation, where fumarate is used as electron donor, electron acceptor and carbon source. This metabolism was analyzed by model simulations and the calculated yields closely matched the actual experimental results. Examination of the genome suggested that *R. ferrireducens* was able to utilize citrate. A gene encoding CitT is 44% identical to a citrate transporter from *E. coli* and is located in a cluster of genes also associated with citrate metabolism. Experimental work confirmed that *R. ferrireducens* can indeed grow on citrate as electron and carbon source with either Fe(III) or nitrate as electron acceptor. The *in silico* model predicted that *R. ferrireducens* could completely oxidize citrate with Fe(III) as electron acceptor and suggested four other pathways of citrate oxidation that also produce acetate and/or succinate under electron acceptor limiting conditions. No fermentative growth was observed in *R. ferrireducens*. Fermentative growth on glucose was simulated with the *in silico* model described above and the result confirmed the experimental observation. Detailed analysis of the metabolic network suggested that the inability of fermentative growth of *R. ferrireducens* is likely due to 1) an inability to convert acetyl-CoA to ethanol, which is in turn consistent with its inability to utilize ethanol; 2) the requirement of ubiquinol in synthesis of biomass components and NADH dehydrogenase as the only carrier able to transfer electrons from NADH to ubiquinone. Thus, a terminal electron acceptor is required for cell growth.

The model proved to be a useful aid in determining which substrates might best support growth for specific applications. For example, growth on eight representative electron donors was simulated under donor-limiting conditions with Fe(III) as electron acceptor. Acetate and glycolate resulted

in the lowest biomass yield, whereas cellobiose produced the highest. Glucose yielded higher biomass than citrate or benzoate.

Genome analyses further suggested that *R. ferrireducens* possesses strategies to manage a number of environmental challenges including aromatic compound metabolism and stresses due to heavy metals and oxidation. *R. ferrireducens* can also cope with nutrient limitation by synthesizing polyhydroxyalkanoates. These storage molecules have industrial and medical interest due to their properties as thermoplastics and elastomers.

This multi-approach study has provided new insights into the remarkable metabolic versatility of *R. ferrireducens* and its potential practical applications. The genome scale model of *R. ferrireducens* will be an important tool for further analysis of the ecology of anaerobic subsurface environments and for the optimization of the unique current-producing capabilities of this organism.

GTL

## Adaptive Evolution of *Geobacter sulfurreducens* under Likely Subsurface Bioremediation Conditions Revealed with Genome Resequencing

Zarath M. Summers<sup>1\*</sup> (zsummers@microbio.umass.edu), Pier-Luc Tremblay,<sup>1</sup> Toshiyuki Ueki,<sup>1</sup> Wael Ismail,<sup>1</sup> Jennifer K. Hart,<sup>1</sup> Bernhard Palsson,<sup>2</sup> Christian Barrett,<sup>2</sup> Karsten Zengler,<sup>2</sup> Yuan Gao,<sup>3</sup> Bin Xie,<sup>3</sup> Chen Liang,<sup>3</sup> and **Derek R. Lovley<sup>1</sup>**

<sup>1</sup>Dept. of Microbiology, University of Massachusetts, Amherst, Mass.; <sup>2</sup>University of California San Diego, La Jolla, Calif.; and <sup>3</sup>Virginia Commonwealth University, Richmond, Va.

**Project Goals:** The goal of this project is to provide computational tools to predicatively 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.

In situ bioremediation of subsurface environments contaminated with organic and/or metal pollutants typically relies on the activity of members of the natural subsurface microbial community. The conditions imposed on that community in order to optimize bioremediation can present circumstances that the subsurface organisms have never previously experienced. Therefore, there may be substantial selective pressure on subsurface microorganisms to change their physiological properties by accumulating beneficial mutations during the implementation of subsurface bioremediation. If so, this could result in dramatic improvements in bioremediation rates and/or effectiveness over time. Thus, documenting which mutations accumulate could aid in better understanding the physiology and ecology of complex bioremediation reactions as well as the interactions of the bioremediating microorganisms with their environment.

For example, *Geobacter* species have been shown to play an important role in the degradation of a diversity of organic contaminants in subsurface environments and are the predominant microorganisms during removal of uranium and vanadium from contaminated groundwater during in situ bioremediation of these contaminants with added acetate. The growth of *Geobacter* species in uncontaminated subsurface environments is generally limited by low availability of electron donors and the difficulties in accessing insoluble Fe(III) oxides, their primary electron acceptor. Therefore, there is little selective pressure for rapid metabolism and growth. This contrasts with the abundance of electron donor when organics are added to the subsurface to stimulate dissimilatory metal reduction or that result from contamination of the subsurface with organic compounds. Furthermore, some of the organic compounds that may be added to the subsurface in some bioremediation strategies are not normally important electron donors in the subsurface and their sudden availability can significantly change the rules of substrate competition.

As an initial evaluation of the potential for adaptive evolution during bioremediation involving *Geobacter* species several studies simulating selective pressures that might be imposed during bioremediation were conducted with *Geobacter sulfurreducens*. For example, slow-release lactate polymers represent a simple method of providing electron donor in the subsurface that may be preferable to the current practice of continually pumping acetate into the subsurface. *G. sulfurreducens* grows poorly on lactate, but the genome-scale metabolic model for this organism predicted that it could potentially grow faster on lactate than on its common electron donor, acetate. Five parallel strains of *G. sulfurreducens* were continually transferred in medium with lactate as the sole electron donor and carbon source. Over a two-year period the doubling time of the lactate cultures decreased from 22 hours to 5 hours, yielding strains that could grow as fast on lactate as on as the ancestral culture grew on acetate.

The lactate-adapted strains were examined with a combination of resequencing chip technology and Illumina sequencing of genomic DNA. All of the lactate-adapted strains had single-base pair substitutions in the gene, GSU0514. Further

investigation, including DNA binding and footprint assays, revealed that GSU0514 encodes a transcription factor that, when expressed, represses transcription of succinyl-CoA synthetase. Succinyl-CoA synthetase is an important TCA cycle enzyme that converts succinyl-CoA to succinate with the generation of ATP. Succinyl-CoA synthetase activity is not required for growth on acetate because succinyl-CoA can be converted to succinate by acetylCoA-transferase, which simultaneously activates acetate to acetyl-CoA for oxidation in the TCA cycle. However, during growth on lactate, there is not sufficient free acetate to accept CoA from succinyl-CoA and thus succinyl-CoA synthetase activity is required. Microarray analysis of gene expression in one of the five adapted strains revealed that expression of the succinyl-CoA synthetase was upregulated with transcript abundance 6.5 fold higher than in wild-type cells. Furthermore, succinyl-CoA synthetase enzymatic activity could not be detected in wild-type cells but was present in the adapted strains.

These results suggested that the mutations in GSU0514 limited the binding capacity of this transcriptional regulator. When GSU0514 was deleted from wild-type cells the knock-out mutant had enhanced growth on lactate, but slightly slower than the adapted strain. Knocking in a copy of GSU0514 that had the same single base-pair change found in one of the adapted strains resulted in a wild-type strain that grew as well on lactate as the adapted strain. The difference in growth on lactate between removing the gene and slightly altering its sequence suggests that this transcription factor influences the expression of a diversity of genes. This is consistent with the finding that over 100 additional genes are differentially expressed in the adapted strain at more than a two-fold difference, including response regulators.

The greater availability of electron donors during bioremediation, alleviates the limitation of growth by electron donors, establishing selective pressure for strains of *Geobacter* that can utilize Fe(III) oxides more rapidly. Over a two-year period *G. sulfurreducens* was continually transferred as rapidly as possible in medium with Fe(III) oxide as the sole electron acceptor and unlimited electron donor availability. The adapted strains now reduce the same amount of Fe(III) in seven days that the ancestral, wild-type strain took 40 days to reduce. Resequencing the genome of one of the adapted strains revealed a single base pair change in the regulatory region 99 bases up stream of the gene that encodes the outer-membrane, c-cytochrome, OmcR. Although microarray data is not yet available, quantitative PCR analysis demonstrated that transcripts of omcR were 19-fold more abundant in the adapted strain than in ancestral strain. These results are of interest because they implicate OmcR in Fe(III) oxide reduction. The appropriate knock-out and knock-in mutations in omcR and the putative regulatory region are currently under construction.

These studies demonstrate that single base pair mutations in regulatory regions of the *G. sulfurreducens* genome that accumulate as the result of new selective pressures that *Geobacter* species are likely to experience during in situ bioremediation

of subsurface environments can dramatically impact on the range of electron donors utilized and the capacity for Fe(III) oxide reduction. Thus, long-term bioremediation of subsurface environments may be viewed as large-scale adaptive evolution experiments and the possibility for improvements in microbial performance as a consequence of this evolution should be considered when attempting to predict the outcome and performance of different bioremediation strategies.

—  
GTL

## Novel Mechanisms Regulating the Expression of Genes in *Geobacter* species Important for Metal Reduction, Electricity Production, and Growth in the Subsurface

Toshiyuki Ueki\* (tueki@microbio.umass.edu),  
Byoung-Chan Kim, and Derek R. Lovley (dlovley@microbio.umass.edu)

Dept. of Microbiology, University of Massachusetts,  
Amherst, Mass.

**Project Goals: 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.**

Genome-scale *in silico* models of *Geobacter* species are currently being developed to predict the activities of *Geobacter* species during *in situ* uranium bioremediation and electricity-harvesting from waste organic matter and renewable biomass. To implement regulatory modules in these models, we have been investigating molecular mechanisms regulating gene expression in *Geobacter* species under various environmentally relevant conditions.

For example, the genomes of *Geobacter* species encode an unusually large number of homologues of two-component systems. This may reflect the need for *Geobacter* species to adapt to a wide range of conditions in subsurface environments. We discovered a novel two-component system in *G.*

*sulfurreducens*, designated GsuTCS1. The sensor histidine kinase component has a novel sensor domain, which contains two putative  $c$ -type heme binding motifs. The response regulator is in the RpoN-dependent enhancer-binding protein family. Our previous studies suggested that GsuTCS1 was involved in redox sensing, Fe(III) reduction, and biofilm formation. Biofilm formation is an important process for optimal current production in microbial fuel cells. To better understand the function of GsuTCS1, the transcriptome of a strain that overproduces GsuTCS1 was compared with the transcriptome in the wild-type cells. Genes with higher transcript abundance in the overproducing strain encode membrane proteins such as a cytochrome  $c$ , a transporter, and a porin. Biochemical analyses further identified direct target genes for GsuTCS1, most of which encode hypothetical proteins with unknown function, but these hypothetical proteins were predicted to have a signal peptide, suggesting that they function in the cellular membrane. These results suggest that GsuTCS1 regulates genes involved in modifying membrane structures during redox sensing, Fe(III) reduction, and electricity production.

*Geobacter* species have been shown to fix atmospheric nitrogen during *in situ* uranium bioremediation and this ability seems to be a key to the predominance of *Geobacter* species in a variety of subsurface environments. We have identified novel regulatory cascades controlling gene expression during nitrogen fixation in *Geobacter* species by a systematic analysis integrated with functional and comparative genomics in combination with biochemical and genetic methods. Unlike regulatory mechanisms known in other nitrogen-fixing microorganisms, the nitrogen-fixation gene regulation in *Geobacter* species was controlled by two two-component His-Asp phosphorelay systems. One of the systems, GnfL/GnfM, was the master regulator that activated transcription of the majority of nitrogen-fixation genes. In addition, the GnfL/GnfM system repressed the gene encoding glutamate dehydrogenase during nitrogen fixation. The GnfL/GnfM system appeared to be essential for growth even in the presence of fixed nitrogen, as a deletion mutant of the GnfL/GnfM system could not be obtained. Overexpression of the GnfL/GnfM system resulted in induction of the nitrogen-fixation genes and repression of the glutamate dehydrogenase gene in the presence of fixed nitrogen, suggesting that the GnfL/GnfM system regulates the nitrogen-fixation genes as well as the glutamate dehydrogenase gene. The other system, GnfK/GnfR, which was directly activated by the GnfL/GnfM system, was shown to be essential for growth in the absence of fixed nitrogen. The amino acid sequence analysis of GnfR indicated that GnfR was an antiterminator. Deletion of *gnfK* and *gnfR* resulted in premature transcription termination of a subset of the nitrogen-fixation genes, whose transcription is activated by the GnfL/GnfM system and which have transcription termination signals in their promoter regions. These results further demonstrate that the GnfK/GnfR system controls by transcription antitermination the expression of the subset of the nitrogen-fixation genes. This study provides a new paradigm to nitrogen-fixation gene regulation.

The TCA cycle is important in *Geobacter* species because it represents the main pathway for the generation of energy and serves to synthesize precursor metabolites. For instance, *in situ* transcript levels of the gene for the citrate synthase, a key enzyme in the TCA cycle, can serve as a biomarker to monitor the metabolic activities of *Geobacter* species during uranium bioremediation and electricity-harvesting. Therefore, the regulation of genes involved in biosynthesis and energy generation was investigated. We identified a transcriptional repressor, designated HgtR, which regulates the expression of genes involved in biosynthesis and energy generation. HgtR is a novel transcription factor, whose homologues are only found in *Geobacter* species. The expression of the *hgtR* gene increased during growth with hydrogen as the electron donor. A strain in which *hgtR* was deleted could not grow on hydrogen. The deletion or overexpression of *hgtR* resulted in activation or repression, respectively, of genes involved in biosynthesis and energy generation. These results suggest that HgtR is a global transcriptional repressor that regulates the genes involved in biosynthesis and energy generation in *Geobacter* species. Moreover, it appeared likely that *hgtR* expression was regulated by the enhancer-binding protein containing a hydrogenase domain. This study paves the way to better understanding of genome-scale metabolic gene regulation in *Geobacter* species.

Riboswitches, noncoding RNA elements found in the untranslated region of mRNA, sense and bind cellular metabolites to control gene expression. *Geobacter* species have been predicted to contain riboswitches sensing a variety of cellular metabolites. For instance, cyclic di-GMP, which functions as a second messenger to regulate diverse physiological processes in bacteria, might be sensed by riboswitches in *Geobacter* species, as homologues of the riboswitch that is known to sense cyclic di-GMP in other bacteria were identified in the genomes of *Geobacter* species. *G. uraniireducens* has the largest number of cyclic di-GMP riboswitch homologues among bacteria whose genomes have been sequenced. In *G. sulfurreducens*, genes known to be differentially regulated during metal reduction and electricity production, such as *omcS* and *omcT*, were found to contain a cyclic di-GMP riboswitch signature in their noncoding region of mRNA. The 5' untranslated region of *omcS* mRNA was shown to be critical for *omcS* expression. These suggest that riboswitches play a role in the regulation of key genes involved in metal reduction and current production.

This increased understanding of regulation of process central to the growth of *Geobacter* species in subsurface environments and the optimization of groundwater bioremediation and electricity production not only contributes to a better understanding of the physiology and ecology of *Geobacter* species, but also provides important information for developing new bioremediation and electricity harvesting strategies.

## Nitrogen Metabolism in the *Geobacteraceae*: The Role of RpoN

Karsten Zengler<sup>1\*</sup> (kzengler@ucsd.edu), Yu Qiu,<sup>1</sup> Jessica De Ingeniis,<sup>2</sup> Derek Lovley<sup>3</sup> (dlovley@microbio.umass.edu), and Bernhard Ø. Palsson<sup>1</sup>

<sup>1</sup>Dept. of Bioengineering, University of California San Diego, La Jolla, Calif.; <sup>2</sup>Burnham Institute for Medical Research, La Jolla, Calif.; and <sup>3</sup>Dept. of Microbiology, University of Massachusetts, Amherst, Mass.

**Project Goals:** The goal of this project is to provide computational tools to predicatively 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.

Nitrogen assimilation is essential for the production of proteins, nucleic acids and cell wall components in prokaryotes and is therefore under stringent control. To study nitrogen assimilation and its control we used genome-wide binding profiles of RNA polymerase (RNAP) and major sigma factors  $\sigma^{70}$  and  $\sigma^N$  in *G. sulfurreducens* under various conditions by ChIP-chip analysis in combination with gene expression profiles using high-density tiling arrays. Over 150 binding sites for  $\sigma^N$  (RpoN) were experimentally identified in *G. sulfurreducens*. All key genes for N-assimilation, nitrogenase, glutamate dehydrogenase, glutamine synthetase and glutamate synthase, were controlled by RpoN, proving that RpoN is the global regulator for N-metabolism in *G. sulfurreducens*. Besides the important role in controlling nitrogen metabolism, RpoN was identified to be an essential sigma factor controlling various other aspects of metabolism in *G. sulfurreducens*. For examples genes responsible for ferrous iron transport were under  $\sigma^N$  (and Fur) control. These genes were only expressed under fumarate reducing conditions in fast growing cells (NH<sub>3</sub> assimilating) but not in slow growing cells (N<sub>2</sub> assimilating). These genes were also not expressed under Fe(III)-reducing conditions (slow growing cells).

Besides ammonium and nitrogen gas we determined that *G. sulfurreducens*, *G. uraniireducens* and *G. metallireducens*

were all able to assimilate glutamine as sole N-source although cells grow significantly slower with glutamine as N-source. Gene expression profiles of *G. sulfurreducens* demonstrated that a variety of genes (e.g. *nif* genes) were up-regulated while growing with N<sub>2</sub> but also with glutamine as compared to NH<sub>3</sub>. Key genes involved in nitrogen assimilation, glutamate dehydrogenase (GDH), as well as glutamine synthetase (GS) and glutamate synthase (GOGAT) were differentially expressed during growth with different nitrogen sources. Cells grown with ammonium as nitrogen source had a higher expression level of GDH while levels of GS and GOGAT were significantly down-regulated compared to growth with N<sub>2</sub> and glutamine, reflecting gene expression profiles for excess of nitrogen and N-limiting condition.

*Geobacter* strains responded to nitrogen limitation in a similar way as described for *Escherichia coli* and *Salmonella typhimurium*. However, in these bacteria, N-limitation is being sensed by the internal glutamine pool (a high glu/gln ratio represents N-limiting conditions) therefore GS and GOGAT expression is up-regulated. *G. sulfurreducens* did respond to N-limitation with up-regulation of GS and GOGAT but the organism did not perceive external N-limitation as internal glutamine limitation. The intracellular glu/gln ratio was in fact an order of magnitude lower for glutamine grown cells as for cells grown with ammonium. In addition, expression of both genes (GlnE and GlnD) that negatively control GS activity directly by adenylation and uridylation was significantly down-regulated in glutamine grown cells. We therefore conclude that members of the *Geobacteraceae* perceive N-limitation differently as shown before for other microorganisms. This finding is in agreement with the proteinaceous sparse environments these organisms are typically found in.

—  
GTL

## Structures and Redox Potentials of Periplasmic Cytochromes from *Geobacter sulfurreducens*

P. Raj Pokkuluri<sup>1\*</sup> (rajp@anl.gov), Yuri Y. Londer,<sup>1</sup> Norma E.C. Duke,<sup>1</sup> Miguel Pessanha,<sup>2</sup> Leonor Morgado,<sup>2</sup> Ana P. Fernandes,<sup>2</sup> Carlos A. Salgueiro,<sup>2</sup> and Marianne Schiffer<sup>1</sup> (mschiffer@anl.gov)

<sup>1</sup>Biosciences Division, Argonne National Laboratory, Argonne, Ill.; and <sup>2</sup>Requimte, CQFB, Dep. Quimica, FCT-UNL, Caparica, 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 structures, and the functional interpretation of the structures, such as active sites and surface properties.**

The genome of *Geobacter sulfurreducens* (Gs) encodes a large number of c-type cytochromes that are expected to function in metal reduction pathways. Of the 111 cytochromes predicted, 91 were identified as produced in one or the other growth conditions tested.<sup>1</sup> Multiheme cytochromes have been implicated in the electron transport chain(s) used by the organism in the reduction of soluble or insoluble metal oxides, or transfer of electrons to electrodes. In addition, the cytochromes may be serving as capacitors (electron-storage sinks) that allow the continued electron flow from the inner membrane to the periplasm in the absence of suitable electron acceptors. Such an electron storage capability will allow the *Geobacter* species to move towards zones where electron acceptors are available, as suggested by Nunez and co-workers.<sup>2</sup>

The periplasmic cytochromes from the c<sub>7</sub> family are required for the reduction of Fe(III) and U(VI) by Gs.<sup>3,4</sup> The periplasmic triheme cytochrome c<sub>7</sub>, PpcA (encoded by GSU0612) was characterized by Lloyd et al., 2003.<sup>5</sup> We expressed PpcA in *E. coli*<sup>6</sup> and determined its three-dimensional structure at 1.45 Å resolution.<sup>7</sup> Further, we identified several sequences homologous to PpcA in the Gs genome.<sup>7</sup> Four of them are of the same size as PpcA; three others are polymers of c<sub>7</sub>-type domains, two of which consist of four domains and one of nine domains that contain a total of 12 and 27 hemes per protein, respectively.

We cloned, expressed, purified, crystallized and determined the structures by X-ray diffraction of all four triheme homologs of PpcA, encoded by GSU0364, GSU0365, GSU1024, and GSU1760. Further we characterized the microscopic reduction potentials of the individual hemes in each protein using NMR and UV-Vis spectroscopic techniques. All the c<sub>7</sub> family cytochromes from Gs have very similar heme core structures. Their overall protein fold is similar with some local structural differences observed. Their structural similarity is highest near heme IV and lowest near heme I. All the proteins have a positively charged surface near heme IV. Such a positively charged surface patch was also observed near the equivalent heme IV in the family of cytochromes c<sub>3</sub>. The region around heme IV in cytochromes c<sub>3</sub> was suggested to be the interaction site of their physiological partner, hydrogenase.<sup>8</sup> By analogy, this region around heme IV in cytochromes c<sub>7</sub> may be the interaction site with their physiological reductase. It can also be speculated that all cytochromes c<sub>7</sub> in Gs interact with the same partner near heme IV as the structure and surface charge are most conserved in this part of their structures. The structures and surface charges have prominent differences near heme I between each of the cytochromes c<sub>7</sub>. Therefore, we suggest that these cytochromes might interact with different partners near heme I or heme III. We determined that the individual heme reduction potentials are different in each protein resulting in different order of oxidation of the hemes in each molecule. This further suggests their functional differences. Studies of mutant Gs strains carrying knockouts of the cytochromes c<sub>7</sub> by Dr. Laurie DiDonato in Prof. Lovley's group displayed different iron reduction rates implying a unique role for each of the proteins.<sup>3</sup>



Multiheme cytochromes that could form protein “nanowires” were identified in Gs and represent a new type of multiheme cytochromes.<sup>7</sup> The sequences of these (two 12-heme GSU1996, GSU0592 and one 27-heme containing GSU2210) proteins suggest that they are formed by domains homologous to triheme cytochromes  $c_7$ . In each domain of the above polymers the heme equivalent to heme IV has His-Met coordination whereas in  $c_7$  all three hemes have bis-His coordination. We have previously determined the structure and measured the macroscopic redox potential of one representative domain (domain C) of a dodecaheme cytochrome (GSU1996).<sup>9</sup> Further, we determined the structures of hexaheme containing domains, AB and CD, and the full four domain protein that could be considered a protein wire. Recently, the microscopic redox properties of the heme groups of domain C were determined using NMR and UV-visible spectroscopies. As observed in cytochromes  $c_7$  family, the hemes in domain C also have different microscopic redox potentials: heme I, -71 mV; heme III, -146 mV; heme IV, -110 mV, which are modulated by strong redox interactions that are dominated by electrostatic effects. The order of oxidation of the hemes is III (His-His coordination) – IV (His-Met) – I (His-His). This result is rather surprising as the His-Met coordinated heme IV does not have the highest potential. It suggests that the polypeptide chain surrounding the hemes plays a dominant role in controlling the individual heme potentials.

This work combines high resolution structural studies and detailed microscopic redox potential characterization of a family of periplasmic triheme proteins and a protein domain. Our results demonstrate that the strong redox interactions between closely packed hemes and their specific interactions with the polypeptide surroundings in structurally similar proteins can extend the range of working heme redox potentials. Such a modulation in heme potentials is likely to have a functional significance in the metabolism of Gs.

## References

1. Ding, Y.-H. R. et al. *Biochim. Biophys. Acta Proteins & Proteomics*, 1764, 1198-1206 (2006).
2. Esteve-Núñez, A. et al. *Environ. Microbiol.* 10, 497-505 (2008).
3. DiDonato, L.N., Lovley, D.R. et al, private communication.
4. Shelobolina E.S. et al. *BMC Microbiology*, 7:16 (2007).
5. Lloyd J.R. et al. *Biochem. J.* 369, 153-161 (2003).
6. Londer Y.Y. et al. *Biochim. Biophys. Acta Bioenergetics*, 1554, 202-211 (2002).
7. Pokkuluri P.R. et al. *Biochemistry*, 43, 849-859 (2004).
8. Yahata N. et al. *Biochemistry*, 45, 1653-1662 (2006).
9. Pokkuluri P.R. et al. *Prot. Sci.*, 13, 1684-1692 (2004).

## Caulobacter Systems Understanding

GTL

### Automatic Segmentation and Structural Study of the Bacterial Cell Wall in Cryo-Electron Tomography

Farshid Moussavi<sup>1\*</sup> (farshid1@stanford.edu), Fernando Amat,<sup>1</sup> Luis R. Comolli,<sup>2</sup> Kenneth H. Downing,<sup>2</sup> Mark Horowitz,<sup>2</sup> and **Harley McAdams**<sup>3</sup> (hmcadams@stanford.edu)

<sup>1</sup>Electrical Engineering, Stanford University, Stanford, Calif.; <sup>2</sup>Life Science Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.; and <sup>3</sup>Dept. of Developmental Biology, Stanford University, Stanford, Calif.

**Project Goals: The objective of this subproject is to develop algorithms to improve and accelerate analysis of electron microscopy tomography images. In particular, we make use of artificial intelligence and statistical image processing algorithms to detect and enhance biological structures within whole bacterial cell images that have high noise levels. This is a component of the overall project thrust to identify and characterize multiprotein complexes in living cells.**

The advent of high throughput tomography is key to structural studies of cellular and subcellular assemblies, and remains an elusive goal. In recent years, there have been tremendous advances in the automatic acquisition of electron microscopy data<sup>1</sup> and generation of tomograms.<sup>2,3</sup> The last remaining bottleneck in the pipeline is the automatic segmentation of cellular structures. Due to severe signal to noise limitations and missing data, established image processing computer vision programs have significant challenges in these datasets. For example, apparent boundaries can be inconsistent, broken, and sometimes just nonexistent due to the missing data wedge. The segmentation of such datasets often involves days of manual effort of an expert.

To address these problems, we have developed a probabilistic framework based on dynamic Bayesian networks (DBN's) that makes significant use of context and shape as well as physical features. A weak shape prior is assumed, and gradually refined as the inference progresses through the 3D volume. So far we have obtained encouraging results on several cryo EM datasets of *Caulobacter crescentus*, achieving automatic segmentation of a membrane in less than 2 hours on a desktop computer. We are extending the model to be more robust to shape changes and different datasets.

One of our first applications of this segmentation is the study of the S-layer structure in its native state. While bacterial S-layers have been studied for over 30 years, most of the studies have been performed on isolated S-layer sheets or proteins.<sup>4,5</sup> We present one of the first structural studies

of S-layer in native state and show some differences with previous results.

We used datasets from cryo-electron tomography (ET) on whole cells of the gram-negative bacterium *Caulobacter crescentus* to obtain quantitative information on the S-layer structure and its interactions with the outer membrane in its native state. We employ pattern recognition and statistical analysis techniques to process efficiently the large volume of low SNR data that results from this method. Using the automatic membrane segmentation described above, we search for the S-layer in a thin volume around the model's estimated surface, by locally maximizing the characteristic S-layer hexagonal signature in the Fourier domain. Due to inevitable noise in the cryo-EM tomograms, we need to model the S-layer as a Markov Random Field to introduce smoothness constraints between neighboring subunits in the surface to minimize false local maxima.

We have identified over 4000 S-layer subunits per tomogram. Visualization reveals that the S-layer structure on the cellular surface of *C. crescentus* CB15 is not as ordered as has been observed in isolated S-layer sheets<sup>5</sup>. Irregularities are present on the surface which break the global<sup>6</sup> symmetry expected in a crystalline structure. We also observe variations on the periodicity of the lattice. We are repeating this analysis on additional *Caulobacter* strains known to have altered S-layer properties.

## References

1. Zheng, S.Q. et al. UCSF tomography: An integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J. Struct. Biol.* 157:138-47, 2007.
2. Lawrence A, Bouwer, J.C., Perkins, G., and Ellisman, M.H.. Transform-based backprojection for volume reconstruction of large format electron microscope tilt series. *J. Struct. Biol.* 154:144-67, 2006.
3. F. Amat, F. Moussavi, L.R. Comolli, G. Elidan, K.H. Downing, M. Horowitz. Markov random field based automatic image alignment for electron tomography. *J. Struct. Biol.* 161:260-275, 2008.
4. Engelhardt, H. Are S-layers exoskeletons? The basic function of protein surface layers revisited. *J. Struct. Biol.* 160:115:124, 2007.
5. Smit, J. et al. The S-layer of *Caulobacter Crescentus*: Three-Dimensional Image Reconstruction and Structure Analysis by Electron Microscopy. *J. Bacteriol.*, 174 (20):6527:6538, 1992.
6. Bartesaghi, A. et al. Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* 162:436:450, 2008.

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" (PI, Harley McAdams, Stanford University).

## Control Mechanisms for Chromosome Orientation and Dynamics in *Caulobacter crescentus*

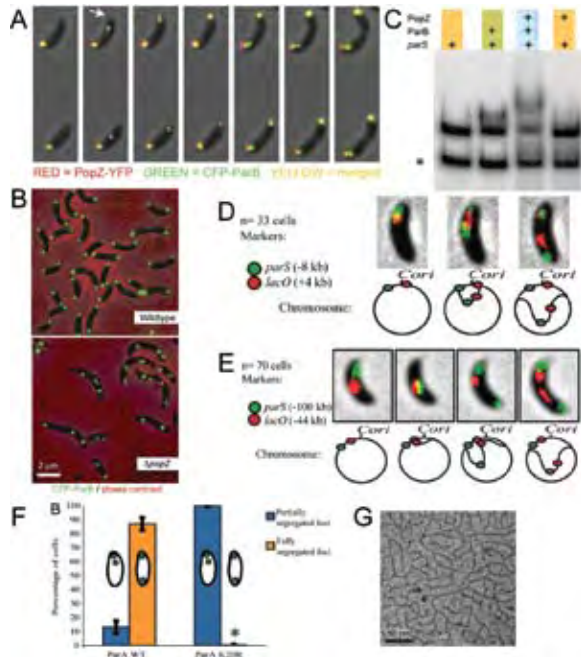
Esteban Toro,<sup>1</sup> Grant R Bowman,<sup>1</sup> Luis R Comolli,<sup>2</sup> Sun-Hae Hong,<sup>1</sup> Jian Zhu,<sup>2</sup> Michael Eckart,<sup>1</sup> Marcelle Koenig,<sup>1</sup> Kenneth H. Downing,<sup>2</sup> W.E. Moerner,<sup>1</sup> Thomas Earnest,<sup>2</sup> **Harley McAdams,<sup>1</sup>** and Lucy Shapiro<sup>1\*</sup> (shapiro@stanford.edu)

<sup>1</sup>Stanford University, Stanford, Calif.; and <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: We seek to develop a global and integrated view of dynamically changing structures associated with cell cycle progression and polar development. Here, we identify and characterize multiprotein complexes associated with chromosome replication and segregation, and elucidate the mechanism by which they carry out large-scale organization within the bacterial cell.**

The application of state-of-the-art cell biology to the bacterial cell has yielded a paradigm shift in our understanding of these tiny cells. Our work has clearly shown that the bacterial cell is highly organized and all cell functions are integrated within the confines of a three dimensional grid. An ongoing and unanswered question is how the bacterial cell segregates its chromosomes. We have shown that in *Caulobacter*, DNA replication and segregation occur simultaneously and that the order of genetic loci on the chromosome reflects the location of these loci in the cell. Prior to the initiation of replication of the single circular chromosome, the origin region of the chromosome, which is 8 kb from the *parS* site, resides at one cell pole and the terminus at the other. Upon replication initiation, a duplicated origin region moves rapidly to the opposite pole. Thus, chromosome organization and segregation in bacteria is precise and regulated. We have identified a novel proline-rich protein, PopZ, as the anchor that tethers separated chromosome origins to the pole. PopZ localizes to the cell pole by a diffusion/capture mechanism and assembles into a filamentous network of high molecular weight oligomers. Further, PopZ interacts directly with ParB, which in turn binds *parS* sequences near to the origin of replication. Thus, at G1, the chromosome is fixed at the cell pole by the PopZ/ParB/*parS* complex. Subsequently, this interaction is released when one copy of the origin region is translocated to the opposite pole. We used chromosomal inversions and *in vivo* time-lapse imaging to show that *parS* is the site of force exertion during segregation, independent of its position in the chromosome. When *parS* is moved farther from the origin, the cell waits for it to be replicated before segregation can begin, and *parS* still forms the leading edge of the translocating chromosome. Also, a mutation in the ATPase domain of ParA halts segregation without affecting replication initiation, suggesting that ParA is the motor that drives ParB/*parS* movement. As the ParB/*parS* complex travels across the cell, PopZ accumulates at the new pole and, upon arrival, tethers

it in place. Chromosome segregation in *Caulobacter* cannot occur unless a dedicated *parS* guiding mechanism initiates movement. Together, the results show that ParB is part of a complex interaction network that controls multiple aspects of chromosome organization, including the translocation of the newly replicated chromosome and its attachment to the cell pole.



**Figure.** *ParB* is part of a complex network that mediates chromosome translocation and docking to the cell pole. (A) After translocation, *ParB* (green) co-localizes with *PopZ* (red) at the cell pole. Two cells were imaged by time-lapse fluorescence microscopy. (B) *ParB* foci (green) are no longer localized to the cell poles in a  $\Delta popZ$  strain. (C) *PopZ* interacts with *ParB* in complex with the *parS* DNA target sequence. A gel shift assay using purified components. The asterisk marks control DNA with no *parS* site. (D-E) Segregation pattern of chromosomal loci in a normal strain (D) and where the order of loci was modified by inversion (E). *Cori* indicates the origin of replication, *parS* indicates the specific *ParB* binding site, and *lacO* is a differentially marked locus. In both cases, *ParB/parS* is the first region to be translocated, regardless of its placement relative to *Cori* or the timing of its replication. (F) A *ParA* mutant disrupts chromosome segregation. After synchronization, cells were left to grow normally (WT) or stimulated to express mutant *ParA* K20R. A cartoon representation of the results is shown with the quantified data (bars). (G) *PopZ* assembles into a filamentous structure *in vitro*. Purified *PopZ* protein was placed on a carbon grid and visualized by transmission electron microscopy.

## Optimizing the Detection of Specific Protein Complexes in *Caulobacter crescentus* at the Electron Microscopic Level

Guido M. Gaietta<sup>1\*</sup> (ggaietta@ncmir.ucsd.edu), Ying Jones,<sup>1</sup> Julie Lee,<sup>1</sup> Masako Terada,<sup>1</sup> Grant Bowman,<sup>2</sup> Monica Schwartz,<sup>2</sup> Lucy Shapiro,<sup>2</sup> and Mark H. Ellisman<sup>1</sup> (mellisman@ucsd.edu)

<sup>1</sup>National Center for Microscopy and Imaging Research, Dept. of Neurosciences, University of California, San Diego, Calif.; and <sup>2</sup>Dept. of Developmental Biology, Stanford University, School of Medicine, Beckman Center, Stanford, Calif.

**Project Goals:** The work conducted by the CRBS/NCMIR team under the umbrella of the DOE-GTL project titled “Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis” aims at generating tools for correlated light and electron microscopic analysis of microbial cells. The current efforts are focusing on (1) improving the preservation of ultrastructure and antigenicity by combining high-pressure freezing, freeze substitution, epoxy fixation and post-embedding labeling; (2) generating tools for automated tracking of macromolecular complexes in tomographic reconstructions of microbial cells.

The detection of protein complexes at the supramolecular level in microbial cells is often hindered by high sensitivity to fixatives and detergents, low preservation of the specimen ultrastructure, limited accessibility of the target protein and reduced antigenicity. To overcome these obstacles, we have developed a method that combines rapid aldehyde fixation, high-pressure freezing, freeze substitution, epoxy fixation (HPF/FS/EF) and post-embedding immunolabeling. Rapid primary aldehyde fixation followed by HPF/FS/EF avoids the bulk specimen shrinkage observed in conventionally prepared samples due to solvent dehydration and provides exceptionally well-preserved ultrastructure and morphology. The specimens embedded in the Araldite/Epon resin (we use a mixture of Araldite and Epon resins during the freeze substitution step and in the final embedding of the sample) are ultrathin sectioned (80-100 nm in thickness) and processed for immunostaining. For antigen retrieval purposes, prior to labeling the ultrathin sections are etched with a freshly made solution of saturated sodium hydroxide in 100% ethanol. The etched sections are then labeled with primary antibodies, gold-conjugated secondary antibodies, stained with uranyl acetate, Sato Lead for contrast purposes and viewed at the microscope.

This optimized labeling method was applied to evaluate the distribution of specific protein complexes related to chromosome organization and replication. Here we targeted lac arrays that were genetically inserted at different locations from the bacterium origin of replication or polar proteins

either natively expressed or in recombinant form fused to a fluorescent reporter and/or an epitope tag (HA).

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.

—  
GTL

## Generating, Implementing and Refining of a Custom Algorithm to Localize Ribosomes in Tomographic Reconstructions of *Caulobacter crescentus*

Rick Giuly,<sup>1</sup> James Obayashi,<sup>1</sup> Ying Jones,<sup>1</sup> Guido M. Gaietta<sup>1\*</sup> (ggaietta@ncmir.ucsd.edu), Lucy Shapiro,<sup>2</sup> and Mark H. Ellisman<sup>1</sup> (mellisman@ucsd.edu)

<sup>1</sup>National Center for Microscopy and Imaging Research, Dept. of Neurosciences, University of California, San Diego, Calif.; and <sup>2</sup>Dept. of Developmental Biology, Stanford University, School of Medicine, Beckman Center, Stanford, Calif.

**Project Goals: The work conducted by the CRBS/NCMIR team under the umbrella of the DOE-GTL project titled “Dynamic spatial organization of multi-protein complexes controlling microbial polar organization, chromosome replication, and cytokinesis” aims at generating tools for correlated light and electron microscopic analysis of microbial cells. The current efforts are focusing on (1) improving the preservation of ultrastructure and antigenicity by combining high-pressure freezing, freeze substitution, epoxy fixation and post-embedding labeling; (2) generating tools for automated tracking of macromolecular complexes in tomographic reconstructions of microbial cells.**

A custom algorithm was generated and used to automatically locate centers of ribosomes in 3D volumetric images collected from a wild type strain of *Caulobacter crescentus*. The automated detection of ribosomes is challenging because (1) the 3D image of each bacterium is large (approximately one gigabyte) and therefore computationally demanding; (2) the identification of ribosomes in the image is not trivial, as they have varying intensity and shape and tend to overlap. The algorithm that we have generated consists of two steps: a **seeding step** and a **dynamic simulation step**. In the **seeding step**, the image is thresholded so that clusters of touching ribosomes tend to produce connected component regions (blobs) of a single gray scale value. For each blob, N particles with radius R are created (at the center of mass of the blob). N is proportional to the volume of the blob so that larger blobs will be seeded with more particles. In this way a group of particle “seeds” is placed at each blob. In the **dynamic simulation step**, the particles are moved so that they tend to settle into darker regions of the blob. The rule for movement of the particles is

inspired by physical interaction between molecules, though highly simplified. When particles are sufficiently close they produce a repulsive force, at medium distance they produce an attractive force, and at large distances no force at all. For computational efficiency, particles within a group only produce force on other particles in their group, not on other groups associated with other blobs. In addition to force that particles exert on each other, there is a force based on the local gradient of the grayscale value in the 3D image. A “gray-scale gradient” force proportional to the magnitude of the gradient is applied to the particle so that it tends to move toward darker (minimum valued) regions. Also, at each time step the particle is displaced randomly by a small distance to help in avoiding the probability of the particle settling to local minimums.

The current iteration of our ribosome tracking software suite (source code available at [cytoseg.googlecode.com](http://cytoseg.googlecode.com)) finds approximately 40% of ribosomes in the *Caulobacter c.* sample, when compared against a human tracked 3D mapping of the ribosomes. Some of the identified factors that contribute to the probability of proper tracking are resolution of the tomogram, as well as size, shape and contrast of ribosomes.

Although the algorithm does not locate all ribosomes, it gives a fully automated method of roughly visualizing the distribution. Also, when combined with manual tracing, the automated method serves to reduce the total time needed for ribosome detection. The program can run and return an output of the XYZ coordinates within hours. Comparatively, the collection of similar numbers of ribosomes by manual tracking requires several days, depending on the experience of the tracker. To improve automatic ribosome detection accuracy, in future work, we plan to develop more sophisticated approaches to segmentation incorporating prior knowledge.

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.

## AMD Community Systems Understanding

GTL

### Whole Community Proteomic Approaches to Decipher Protein Information from Natural Microbial Communities

Robert L. Hettich<sup>1\*</sup> (hettichrl@ornl.gov), Nathan VerBerkmoes,<sup>1</sup> Brian Erickson,<sup>1</sup> Alison Russell,<sup>1</sup> Paul Abraham,<sup>1</sup> Chongle Pan,<sup>1</sup> Manesh Shah,<sup>1</sup> Doug Hyatt,<sup>1</sup> Denise Schmoyer,<sup>1</sup> Paul Wilmes,<sup>2</sup> Ryan Mueller,<sup>2</sup> Vincent Deneff,<sup>2</sup> Steve Singer,<sup>3</sup> Michael Thelen,<sup>3</sup> and **Jillian Banfield**<sup>2</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.;  
<sup>2</sup>University of California, Berkeley, Calif.; and <sup>3</sup>Lawrence Livermore National Laboratory, Berkeley, Calif.

The advent of integrated computational and experimental approaches for microbial genomic interrogation has afforded a unique opportunity to study microbial consortia at the molecular level. This permits a detailed examination of how environmental communities cooperate and compete for resources in natural ecosystems. To this end, we have focused on the development of experimental (LC-MS/MS) and computational approaches to characterize microbial communities in their natural settings for elucidation of how these consortia adapt and respond to their environmental pressures. Because the complexity of many natural microbial communities exceeds the current measurement capabilities of analytical techniques, it is advantageous to develop and demonstrate methodologies with a low complexity environmental microbial consortium. In this respect, the acid mine drainage (AMD) microbial system is ideal. Sufficient biomass is readily accessible to enable molecular level evaluation by a variety of genomic, proteomic, and biochemical techniques. This permits coordination of different analytical measurements on the same samples, thereby providing the ability to integrate the datasets for extraction of biological information. Furthermore, the system is moderately rich in bacterial, archaeal, and viral membership.

Whole community genomics serves as the underlying core for almost all of the subsequent measurements and evaluations of microbial consortia. The depth and quality of genome annotation, including information about bacteria (with strain variation details), archaea, and viruses, is critical for the ensuing proteomic and biochemical measurements. The recent availability of a greatly expanded genome annotation of bacteria and archaeal species in the AMD samples has provided a much richer database from which to mine proteome data. This has permitted not only a deeper level of proteome identification in the newer datasets, but also is being used to re-search older existing datasets to mine additional microbial information, especially from the archaeal members.

Examination of extensive proteome datasets from the AMD system to date has prompted the need for new experimental approaches to increase the dynamic range of proteome measurement, thereby providing a deeper and wider view of the range of protein components important for microbial community structure and function. To this end, we have explored two experimental approaches. The first approach relies on multiple mass-range scanning in the mass spectrometer to enable more extensive measurement metrics on the chromatographic time-scales. While this requires a bit more sample, the resulting datasets provide more detailed fragmentation information from a greater number of peptides in the complex mixtures, and thus provides deeper proteome information. The second approach, conducted in close conjunction with the Banfield UCB research group, relies on isoelectric focusing (IEF) fractionation of intact proteins as an additional separation step for our 2-dimensional LC/LC-MS/MS format. This has afforded not only an additional enrichment step for improving the depth of proteome coverage, but also provides an additional identification metric (i.e., pI correlation) of the proteins in each fraction. Although both of these approaches are still under evaluation in our lab, we have already observed significant improvements in proteome discovery (e.g., >6,000 proteins identified using the IEF fractionation procedure) with these and other related approaches.

In addition to improving the dynamic range of proteome measurement, we have also realized the need to better characterize the range of post-translational modifications (PTM) that impact the function of the biological ecosystem. We have focused our efforts into two specific areas. The first research area is a detailed evaluation of the extracellular proteome of the AMD system, in an effort to more fully characterize the proteins that experience the extreme environmental conditions and likely play critical roles in respiration and other vital biological activities. In particular, we have developed an integrated experimental/computational approach to investigate the range of extracellular proteins that have been processed by signal peptide cleavages. We identified ~500 signal peptide cleaved proteins with high confidence, including numerous proteins of unknown function which were further characterized by Pfam analysis. Spatial profiling of signal peptide protein expression exhibited a notable subset of proteins either conserved or divergent in their expression as a function of growth state maturity or spatial location within the mine. In particular, some cytochromes (including Cyt579) and outer membrane proteins were highly conserved across all growth states, while other putative cytochrome-related proteins were found to be highly variable across the growth states. In conjunction with the Thelen LLNL research group, we have initiated a detailed intact protein characterization of Cytochrome 579 in an effort to understand how post-translational modification and sequence variation of this protein might provide environmental tuning as a function of growth state or spatial location within the mine. The second research area involves evaluation of new computational approaches to extract PTM information from the already existing datasets. In particular, we have evaluated a new computational approach based on sequence tagging to identify a defined set of PTMs

(methylation, acetylation, and oxidation). By examining six different AMD proteome datasets, we found: a) about 20% of the peptides are modified in each sample, b) oxidative damage is much higher in the AMD system than in other bacterial isolates, and is enhanced in early growth stages, and c) flagellin and chemotaxis proteins are highly modified in late growth state biofilms, especially for *Lepto III*. This suggests higher motility in late developmental growth stages.

By expanding the range of samples examined for the AMD system, as well as developing and testing new methods of increasing the dynamic range and modification identities of the proteins interrogated, we have been able to uncover a much deeper level of molecular level information about the structure and functions of natural microbial consortia.

This research sponsored by the U.S. DOE-BER, Genomics:GTL Program. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

GTL

## Insights into the Ecology and Evolution of a Natural Microbial Ecosystem from Acid Mine Drainage using Community Genomics and Proteomics

R.S. Mueller<sup>1\*</sup> (rmueller@berkeley.edu), V.J. Deneff,<sup>1</sup> S.L. Simmons,<sup>1</sup> D. Goltsman,<sup>1</sup> P. Wilmes,<sup>1</sup> C. Belnap,<sup>1</sup> A.P. Yelton,<sup>1</sup> G.J. Dick,<sup>1</sup> B.J. Baker,<sup>1</sup> L. Comolli,<sup>3</sup> B.C. Thomas,<sup>1</sup> B. Dill,<sup>2</sup> L. Hauser,<sup>2</sup> M. Land,<sup>2</sup> S.W. Singer,<sup>3</sup> L. Kalnejais,<sup>1</sup> M. Shah,<sup>2</sup> N.C. VerBerkmoes,<sup>2</sup> M.P. Thelen,<sup>3</sup> R.L. Hettich,<sup>2</sup> and J.F. Banfield<sup>1</sup>

<sup>1</sup>University of California, Berkeley, Calif.; <sup>2</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; and <sup>3</sup>Lawrence Livermore National Laboratory, Livermore, Calif.

**Project Goals: Our goal is to develop community genomic and proteomic methods to study evolutionary processes and ecological patterns of natural microbial systems. This is achieved primarily through analysis of model communities comprised of closely and more distantly related co-evolving organisms growing in an acid mine drainage environment. Community genomics methods that address both sequence variation and binning have allowed for an in-depth examination of population genetic heterogeneity, providing insights into the interplay between mutation, recombination, and transfer of mobile genetic elements and selection. Analyses of many whole community proteogenomic datasets for natural and bioreactor biofilms have enabled identification of key factors that impact organismal function, ecosystem partitioning, and dictate overall ecosystem properties.**

Rigorous mining of the wealth of information produced through “omics” methods promises deeper insights into biological systems. While community genomics provides snapshots of variation from which evolutionary processes can be

inferred and generates hypotheses relating to the function of ecosystems, community proteomics captures metabolic responses *in situ*, enabling hypothesis testing.

Acid mine drainage (AMD)-generating systems are characterized by hot (often >40 C), pH < 2 solutions, enriched in toxic heavy metals [mM] that are released by dissolution of sulfide minerals exposed by mining. The organisms that proliferate in AMD are well suited to cope with these extreme stresses and accelerate sulfide mineral dissolution and acid production as a byproduct of metabolism. The constrained biological diversity of this environment, and the intimate relationship between biology and chemistry, makes these ideal model ecosystems. Using the AMD system, our labs seek to develop combined cultivation-independent genomics and proteomics methods for study of microbial communities.

We reconstructed near complete genomes for the principal bacterial and archaeal members of two biofilms collected from the air-AMD interface within the Richmond mine (Iron Mountain, CA). These include typically dominant *Leptospirillum* Group II (LII) bacteria, the lower abundance and perhaps keystone member, *Leptospirillum* Group III (LIII), and Archaeal members from the *Thermoplasmatales* group and ARMAN group. Comparisons of sequences from individuals that comprise natural populations and comparative genomic analyses involving composite sequences from more distantly related but co-evolving species enable novel insights into evolutionary processes.

The fine-scale heterogeneity within LII populations was examined genome-wide to investigate intra-species evolutionary processes. The LII population is dominated by one sequence type, yet we detected evidence for relatively abundant recombining variants (>99.5% sequence identity) at multiple loci, and a few rare variants. Population genetic analyses of single nucleotide polymorphisms indicate variation between closely related strains is not maintained by positive selection, suggesting that these regions do not represent adaptive differences between strains. Thus, the most likely explanation for the observed patterns of polymorphism is divergence of ancestral strains due to geographic isolation, followed by mixing and subsequent recombination. [Simmons et al. 2008]

In order to infer possible niche specialization between community members from the same genus, we compared the genomes of LII and LIII. Both have genes for community-essential functions, including carbon fixation and synthesis of biofilm structural components, but LII is better equipped to deal with osmotic challenges. In contrast, LIII has the genes to fix nitrogen, has more chemosensing pathways, and expresses more motility proteins than LII, consistent with their observed localization within the biofilm interior, where steep chemical gradients are likely.

Further evidence for niche specialization between organisms was obtained through the examination of 28 whole community proteomes obtained from the AMD system over a four-year period. We previously reported the genomes of two LII species, the UBA and 5way types,

whose orthologous proteins differ by 5% sequence identity. Comparative genomic and proteomic analysis revealed both the type-specific expression of orthologous genes as well as the expression of a limited subset of genes unique to each type. The most significant differentiating signal originated from genes involved in cobalamin biosynthesis and cobalamin- and S-adenosyl methionine-dependent anabolic methylation reactions. These, as well as glycine cleavage complex proteins, which generate a C<sub>1</sub> and ammonium, were more abundant in the 5way type, indicating that it might obtain a competitive edge by scavenging carbon sources in late developmental stage biofilms where it is most abundant. [Denef et al. 2008]

Variations in protein abundances may reflect changes in organismal activity in response to fluctuations in biotic or geochemical conditions. We used statistical methods adapted for proteomics data to determine how the community proteome expression patterns correlate with environmental factors. It was found that low abundance members of the microbial population show a strong response to fluctuations of abiotic factors, whereas the overall expression of the dominant member, LII, is less responsive. For example, the protein abundances of many *Thermoplasmatales* members and temperature correlate positively, but the activities of individual members are also affected by additional factors (i.e. *G-plasma* with [SO<sub>4</sub><sup>2-</sup>], *A-plasma* with pH and [Ca]). Using this proteomics approach, we are able to propose niches of community members and their fine-scale relationships with specific environmental determinants, which would not be possible using comparative genomics alone. Another observation from this work is that changes in the community composition strongly correlate with overall LII protein expression. For example, the progression from homogeneous, LII-dominated biofilms to high developmental stage films containing many other organisms coincides with a shift in the LII proteome from one biased toward ribosomal proteins and those involved in stress responses to one biased toward enzymes that produce nucleotides, amino acids and carbohydrates. Thus, the global metabolism of the most abundant organism of these biofilms is dependent on inter-species exchanges and competition, and highly resistant to changes in the prevailing environmental conditions.

These results, coupled with detailed microscopic analysis of intact biofilms from the mine showing layered microbial assemblages [Wilmes et al. 2008], inform an overarching model of ecosystem function and evolution within these microbial communities. As LII is the initial colonist of this community, due to its ability to grow alone in the extreme environment and fix carbon, establishes a stable environment for the propagation of lower abundance organisms. These organisms may perform overlapping roles with the ecosystem, but appear to be adapted to well-defined environmental conditions. Functional capacity and adaptations to differing conditions are driven by genetic heterogeneity and exchange between close relatives, in addition to changes in the regulation of key metabolic pathways in response to the external environment.

Funding provided by DOE Genomics:GTL Program grant number DE-FG02-05ER64134

## Community Proteogenomic Analysis of Virus-Host Interactions in a Natural System

Christine Sun<sup>1\*</sup> (christine\_sun@berkeley.edu), Christopher Belnap,<sup>1</sup> Brian Thomas,<sup>1</sup> Kimberly Pause,<sup>2</sup> Paul Wilmes,<sup>1</sup> Daniela Goltsman,<sup>1</sup> Chongle Pan,<sup>3</sup> Anders Andersson,<sup>1</sup> Robert Hettich,<sup>3</sup> Nathan VerBerkmoes,<sup>3</sup> Mya Breitbart,<sup>2</sup> and Jill Banfield<sup>1</sup>

<sup>1</sup>University of California, Berkeley, Calif.; <sup>2</sup>University of South Florida, Saint Petersburg, Fla.; and <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.

**Project Goals:** The aim of this project is to develop integrated genomic and proteomic (proteogenomic) methods to study virus-host dynamics in bioreactor-grown and natural microbial biofilm communities. We will utilize simultaneous genomic analysis of CRISPRs (the adaptive microbial immune system) and viruses to examine the co-evolution and dynamics of viruses and their microbial hosts. Due to virus-host specificity, viruses can change the microbial community structure by enriching for resistant strains. Proteomics will be used to examine metabolic interplay via monitoring of protein production before, during, and after viral infection, while genomic analysis of CRISPRs will be used to correlate these findings with changes in community structure. Because viral predation can undermine microbially-based technologies such as microbial fuel cells, bioethanol production, and environmental remediation, methods to study microbe-viral interactions are widely relevant to DOE missions.

Despite their tiny size, viruses represent the most abundant entities on Earth. Bacteriophages and archaeal viruses (here collectively referred to as viruses) have the ability to shape the composition and functionality of microbial communities through predation of their hosts and lateral gene transfer. Since viruses can play central roles in microbial evolution, it is important to understand the dynamic interactions between viruses and their microbial hosts. Yet, few studies have examined virus-host dynamics and the diversity of viruses in natural populations. While various metagenomic studies have allowed glimpses into functional gene diversity in viral populations, no cultivation-independent method has existed to link a particular virus with its hosts prior to the identification of CRISPRs (clustered regularly interspaced short palindromic repeats) as adaptive, rapidly evolving microbial immune systems. Simultaneous genomic analysis of the CRISPR loci of natural bacterial and archaeal populations and the virus populations that they target provides a new route to study virus-host interaction dynamics [Tyson et al. 2007; Andersson et al. 2008]. The relatively low diversity microbial community in acid mine drainage (AMD) from Iron Mountain, CA, allows the opportunity to comprehensively examine the dynamics between CRISPR regions and viruses. The findings from this system are likely generally applicable and directly relevant to the DOE mission through the implications for the maintenance of stable

biotechnologies, including bioremediation and industrial bioenergy production.

A CRISPR locus consists of repeats separated by short unique spacer sequences that originate from viral genomes and other mobile elements. The *cas* (CRISPR-associated) genes, which are found adjacent to CRISPR loci, encode proteins that are key to the functioning of the immune system. Microbial immunity to viruses derives from having a CRISPR spacer sequence exactly matching a viral sequence [see Barrangou *et al.* 2007]. Since spacers incorporate into the CRISPR locus in one direction, the CRISPR region provides a record of recent viral exposure. We have shown that CRISPR loci within microbial genomes assembled from AMD community genomic datasets are extremely dynamic genomic regions that display evidence of extensive loss and gain of spacers. Using 454 FLX sequencing of CRISPR loci from two *Leptospirillum* group II bacterial populations, we recovered 521,203 total spacer sequences, of which 16,144 were unique. Rarefaction curves built from these CRISPR spacers show no approach to saturation. The diversity in spacer sequences within host populations is consistent with evolution of the spacer complement on microbial generational time scales.

The viral-derived spacer sequences have allowed us to identify previously unassigned sequences in AMD community genomic datasets as viral and to assemble the viral sequences into 5 near-complete genomes (AMDV1-AMDV5), enabling unprecedented analysis of heterogeneity within natural virus populations [Andersson *et al.* 2008 and unpublished]. Recombination is apparent within all viral populations and in one population (AMDV2) it is sufficiently extensive to serve as a means of avoiding the short sequences of CRISPR spacers.

Quantitative proteomic analyses of replicate, laboratory-grown communities suggest a stochastic nature of viral infection within the AMD biofilms. Three replicate community biofilm samples each contained different viral types (AMDV1 vs. AMDV3 vs. AMDV4), with few to no proteins identified from the other AMD viral types. Given that communities were grown from the same starting inoculum, we infer that viral blooms can be unpredictable and spatially heterogeneous. This may be a consequence of rapid evolution within the “clouds” of host resistance and viral sequence diversity. There is also evidence for spatial and temporal heterogeneity in the host response for *Leptospirillum* group II in these laboratory-grown biofilms. Five of the Cas proteins from *Leptospirillum* group II were ~60 times more abundant in biofilm-associated cells than planktonic cells, yet no such distinction was observed in a replicate community grown within a separate laboratory reactor. This, and related findings for natural communities, indicate strong regulation of the CRISPR/*cas* system. Heterogeneity in viral populations is also apparent at larger scales within the field site. For example, distinct AMDV1 sequences are only found in certain community genomic datasets and identification of Cas proteins from host populations is quite variable.

Aligning CRISPR spacer sequences to the viral genomes reveals information about the still poorly understood

CRISPR mechanism. For example, 1,235 unique spacers derived from 454 sequencing matched AMDV1, the bacteriophage with *Leptospirillum* group II and III as hosts. Different spacers target different variants of the same viral type. In some cases, these spacers overlap with each other and even target different variants in the same genomic region. Also, plasmids have been found to contain CRISPR loci. Plasmid populations reconstructed from AMD community genomic datasets have CRISPRs with spacers that target the *cas* sequences of other plasmid CRISPRs, presumably to silence that plasmid’s immune system. This illustrates a previously unknown phenomenon: CRISPRs may be involved in plasmid versus plasmid competition.

All evidence from the AMD system indicates that both CRISPRs and viruses are co-evolving rapidly, in an “arms-race” that requires continual acquisition of new spacers to target viral sequences that have been modified by mutation and sequence shuffling. The features of this emerging view of virus-host interactions are consistent with findings published from parallel pure culture laboratory studies conducted by other groups [Barrangou, Horvath and others]. We infer that the dynamic balance in infectivity and resistance is a general feature of virus-host interactions, a result that underlines the importance of viruses in shaping the structure and functioning of microorganisms in ecosystems of both natural and biotechnological importance.

Funding provided by DOE Genomics:GTL Program grant number DE-FG02-07ER64505

GTL

## Shotgun Proteomics and *De Novo* Sequencing for the Detection of Viral Signatures in Natural Microbial Communities

Nathan VerBerkmoes<sup>1\*</sup> (nve@ornl.gov), Chongle Pan,<sup>1</sup> Brian Dill,<sup>1</sup> Patricia Carey,<sup>1</sup> Jacque Young,<sup>1</sup> Doug Hyatt,<sup>1</sup> Manesh Shah,<sup>1</sup> Kim Pause,<sup>2</sup> Christine Sun,<sup>3</sup> Brian Thomas,<sup>3</sup> Paul Wilmes,<sup>3</sup> Robert L. Hettich,<sup>1</sup> Mya Breitbart,<sup>2</sup> and **Jillian Banfield**<sup>3</sup> (jbanfield@berkeley.edu)

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; <sup>2</sup>University of South Florida, St. Petersburg, Fla.; and <sup>3</sup>University of California, Berkeley, Calif.

**Project Goals: The aim of this project is to develop integrated genomic and proteomic (proteogenomic) methods to study virus-host dynamics in bioreactor-grown and natural microbial biofilm communities. We will utilize simultaneous genomic analysis of CRISPRs (the adaptive microbial immune system) and viruses to examine the co-evolution and dynamics of viruses and their microbial hosts. Due to virus-host specificity, viruses can change the microbial community structure by enriching for resistant strains. Proteomics will be used to examine metabolic interplay via monitoring of protein production before,**



during, and after viral infection, while genomic analysis of CRISPRs will be used to correlate these findings with changes in community structure. Because viral predation can undermine microbially-based technologies such as microbial fuel cells, bioethanol production, and environmental remediation, methods to study microbe-viral interactions are widely relevant to DOE missions.

Microorganisms comprise the majority of extant life forms and play key roles in a wide variety of health and environmental processes, yet little is known about the nature and driving forces of their diversification. Although the roles of viruses in microbial evolution are widely recognized, neither the details of viral-microbial interactions nor the impact of virus on microbial community structure are well understood. Community genomic and proteomic (proteogenomic) methods have been established for analyzing the roles and activities of uncultivated bacteria and archaea in natural multi-species consortia (Ram, *Science* 2005, Lo *Nature* 2007, Wilmes, *ISME*, 2008, VerBerkmoes, *ISME*, 2008). Notably lacking are methods for monitoring viral activity in communities, tracking virus predation, and determining the consequences of viral predation for ecosystem structure. Our work focuses on acid mine drainage (AMD) biofilms because of their relative simplicity and established utility as a model system for development of methods for cultivation-independent analyses. Our goal is to develop proteomics and informatic techniques to characterize the microbial response to viral attacks as well as to identify the viral signatures in natural microbial communities.

One of our first goals was to examine the distribution and spectral evidence of viral proteins and peptides across all AMD biofilm proteomes analyzed to date (this includes over 8 full proteomes (multiple fraction per proteome) and 40 sub-proteomes (only whole cell fractions)). All proteomes were re-analyzed with a new database that contained the consensus predicted protein sequences for four different contigs hypothesized to contain phage proteins. *The central hypothesis to this part of the study is there will be conserved viral proteins and peptides from those proteins that can be found in multiple sampling sites.* Furthermore the MS/MS spectra, from these conserved viral peptides, should act as molecular fingerprints for viruses in the AMD system. To support the hypothesis it will be necessary to find unique peptides, with high quality spectra, that were reproducible over numerous locations in the mine. The first step of this study was to compile all viral proteins into a central dataset. Using the spectral counts from the database results, those proteins annotated as 'viral' were compiled into a database which includes the annotated name, the location(s) in which they were present, the number of the most abundant spectra and the fraction from which the highest spectral count was found. We were able to find a number of conserved viral peptides across most of the biofilms and manually validated MS/MS spectral evidence for the expression of viral proteins in all biofilms analyzed to date. One of the major questions though is how many viral peptides and proteins are we missing due to strain variation within the viral genomes? One possible solution to this problem is the use of strain resolved protein databases instead of consensus protein databases.

Our current approach was to use a six frame translation of all the viral proteins including all of their possible variants. Currently, we are examining the results from this analysis.

The possibility exist that no matter how many different viral strain resolved proteins, that are in a database, that the virus proteins will evolve differently in different samples; thus making it impossible to identify the viral proteins with the typical database searching algorithms. Alternative advanced analytical and informatic methodologies are needed to identify rapidly evolving viral proteins in complex microbial communities. *De novo* sequencing programs aim to derive complete or partial amino acid sequences from MS/MS scans without complete information from protein sequence databases. Although *de novo* sequencing approaches seem straightforward, they have not been widely applied in proteomics due to low data quality and software limitations. Thus one of our major efforts is to improve *de novo* sequencing methods to specifically be able to identify viral peptides, even if they have changed 1-2 amino acids for every ten amino acids of protein sequence. We have three major subtasks to accomplish this goal: 1) Develop and test improved analytical methods to create better mass spectral data for the algorithms to use; 2) Test all available commercial and public *de novo* and sequencing tagging algorithms; and 3) Develop a new *de novo* sequencing algorithm that can directly use the improved MS data.

*The development of better mass spectral data is critical for de novo sequencing in our view. We believe that the biggest hindrances to high quality de-novo sequencing are low quality MS data.* Our approach is to create datasets using our normal 2d-LC-MS/MS methods that contain three specific analytical data for each peptide sequenced. Our goal for every peptide is to acquire 1) a high resolution parent peptide mass spectra with mass accuracies of 1-5ppm with internal calibrations; 2) a high resolution CID (collisional induced dissociation) spectra 3) a high resolution ETD (electron transfer dissociation) spectra. We have currently established steps 1 and 2 on LTQ-Orbitraps in automated 24 hour runs, we have established linked CID and ETD spectra of the same peptide, at low resolution on LTQs and we feel that as soon as the LTQ-Orbitraps are upgraded with ETD; this should be straightforward to obtain all three data points for all peptides analyzed. The major downside of this approach is duty cycle; in general shotgun proteomics every peptide can be analyzed in 100-200 milliseconds. This integrated approach of linked high resolution full scans, CID, and ETD scans will take at least 1-2 second per peptide. One solution to this problem that we have been exploring is the use of 2d-LC-MS/MS with multiple mass range scanning.

We have compared a number of *de novo* programs. The datasets for comparisons was a standard microbial isolate proteome (*R. palustris*) and an AMD proteome. To mimic unknown strain variation we eliminated the genome of the dominant strain and used a related strain as the template. We evaluated the following: speed, accuracy (total number of amino acids in a string it could find), scoring method (how accurate is the scoring method, do high scores always equal correct identifications, ability to use a diversity of mass

spectral datasets, dynamic range of identifications and general ease of use. After testing we conclude that two of the newer algorithms provide the best results but each had positive and negative aspects to them. PEAKS is a commercial software, it had medium run speed, the best accuracy (*it was the only program we have seen that can get 100% of the peptide correct*), was straightforward to operate and could handle data from multiple MS platforms. DirecTag (available from Vanderbilt University) is a very fast algorithm, but it only provides short tags, its scoring scheme is not as accurate as PEAKS, but it provides tags for many more spectra than PEAKS. From these initial studies it was clear that no one given algorithm, at this point, can provide all the answers or is the best for all potential experiments. Clearly the integration of algorithms and the use of higher quality mass spectral data discussed above will be necessary.

High-resolution tandem mass spectra can now readily be acquired with hybrid instruments, in high-throughput shotgun proteomics workflows, as discussed above. *We have developed a new de novo sequencing algorithm, Vonode, has been developed specifically for such high-resolution tandem mass spectra.* To fully exploit the high mass accuracy, sparse noise, and low background of these spectra, a unique scoring system is used to evaluate sequence tags based mainly on mass accuracy information of fragment ions. Consensus sequence tags were inferred for 11,422 spectra with an average peptide length of 5.5 residues from a total of 40,297 input spectra acquired in a 24-hour proteomics measurement of *R. palustris*. The accuracy of inferred consensus sequence tags was 84%. The performance of Vonode was shown to be superior to the PepNovo v2.0 algorithm, especially in term of the number of *de novo* sequenced spectra in a head to head test. Currently we are developing a fully integrated software package for this software and testing it along side PEAKS and DirecTag.

This research sponsored by the U.S. DOE-BER, Genomics:GTL Program. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

GTL

## Characterization of Viruses from an Acid Mine Drainage System

Kimberly C. Pause<sup>1\*</sup> (kpause@marine.usf.edu), Christine L. Sun,<sup>2</sup> Paul Wilmes,<sup>2</sup> Robert Hettich,<sup>3</sup> Nathan VerBerkmoes,<sup>3</sup> Luis Comolli,<sup>4</sup> Jillian F. Banfield,<sup>2</sup> and Mya Breitbart<sup>1</sup>

<sup>1</sup>University of South Florida, Saint Petersburg, Fla.;

<sup>2</sup>University of California, Berkeley, Calif.; <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, Tenn.; and <sup>4</sup>Lawrence Berkeley National Laboratory, Berkeley, Calif.

**Project Goals: The aim of this project is to develop integrated genomic and proteomic (proteogenomic) methods to study virus-host dynamics in bioreactor-grown and**

**natural microbial biofilm communities. We will utilize simultaneous genomic analysis of CRISPRs (the adaptive microbial immune system) and viruses to examine the co-evolution and dynamics of viruses and their microbial hosts. Due to virus-host specificity, viruses can change the microbial community structure by enriching for resistant strains. Proteomics will be used to examine metabolic interplay via monitoring of protein production before, during, and after viral infection, while genomic analysis of CRISPRs will be used to correlate these findings with changes in community structure. Because viral predation can undermine microbially-based technologies such as microbial fuel cells, bioethanol production, and environmental remediation, methods to study microbe-viral interactions are widely relevant to DOE missions.**

Viruses are the most abundant biological entities on the planet, playing important roles in biogeochemical cycling, horizontal gene transfer, and defining the community composition of their hosts. We are only beginning to understand the identity, diversity, and ecology of viruses in the environment. Community analysis through metagenomic sequencing has proven to be a useful tool for examining viruses in a variety of environments. Studies have revealed diverse environmental viral communities with a high degree of novelty. Previous community analyses of the acid mine drainage (AMD) microbial biofilm communities at our study site (Richmond Mine at Iron Mountain, CA, USA) have demonstrated that this is a low-diversity system that can offer unique insight into virus-host interactions by concomitant genomic sampling of virus populations and clustered regularly interspaced short palindromic repeat (CRISPR) loci (See Sun *et al.* poster).

We have optimized methods to isolate the viral community from two distinct types of AMD biofilm, and used microscopy and metagenomic sequencing to characterize the viral community. A combination of mechanical disruption, filtration, and density-dependent centrifugation successfully enabled purification of viruses from the biofilm. The purity of the fractions was examined using epifluorescence microscopy and PCR amplification of bacterial and archaeal 16S genes. Several different viral morphologies were observed by electron microscopy, including icosahedral capsids typical of bacteriophages, and lemon and rod-shaped morphologies similar to archaeal viruses found in other extreme environments. To verify the presence of viruses in the sample, viral sequences from previously obtained community metagenomic datasets were used to design primers to specifically target AMDV1, a bacteriophage of *Leptospirillum* groups II and III. Sequences generated from these PCR clone libraries exhibited extensive single nucleotide polymorphisms among the AMDV1 virus population.

Metagenomic sequencing was also used to characterize the purified viral community. Preliminary screening demonstrated that approximately 80% of the sequences had similarity to known AMD viral or putative viral sequences, confirming the success of the virus purification method. This may indicate that there is overall low virus diversity at the population level, which is consistent with the low

microbial diversity identified in previous studies in this system. Additionally, our sample was collected in 2007, more than 3 years after the samples were collected from which the genomic database was developed. This suggests that there may be long-term stability of virus populations in the AMD system, making it an ideal place to study viral evolution in a natural setting. An important next step is to examine purified virus DNA from the AMD biofilm in order to compare CRISPR loci to coexisting viruses and identify viruses not targeted by the CRISPR system.

Finally, we are working with *Leptospirillum* group II isolates, originally cultured from the mine for a previous study. Efforts are underway to infect these cultures with viruses isolated from the biofilm samples, with a goal of isolating virus-host systems for laboratory manipulation. We are also working to characterize the vesicles produced by the *Leptospirillum* group II isolates, which are found in culture as well as in the natural environment. Insight gained from these experiments may help in eliminating vesicles and any associated nucleic acids from the purified viral fraction. Additionally, we will examine the role that these vesicles may play in gene transfer.

—  
GTL

## Microbially Mediated Transformation of Metal and Metal Oxide Nanoparticles

**M.J. Doktycz**<sup>1\*</sup> (doktyczmj@ornl.gov), D.A. Pelletier,<sup>1</sup> C.K. McKeown,<sup>1</sup> S.D. Brown,<sup>1</sup> B. Gu,<sup>2</sup> W. Wang,<sup>2</sup> D.P. Allison,<sup>1,3</sup> D.C. Joy,<sup>3,4</sup> and T.J. Phelps<sup>1</sup>

<sup>1</sup>Biosciences Division and <sup>2</sup>Environmental Sciences Division, <sup>3</sup>University of Tennessee, Knoxville, Tenn.; and <sup>4</sup>Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tenn.

**Project Goals:** The goal of the project is to understand the potential fate and transport of engineered materials in the environment. To do so, the interactions between engineered nanoparticles and bacterial systems are being investigated. The technical approach involves three aims. First, the preparation of well-characterized metal and metal oxide nanoparticles is being pursued. Well-defined nanoparticle structures are essential for drawing accurate conclusions on how physical and chemical characteristics influence microbial response. The response of various microbial species to these materials will be assessed by conventional bacterial growth, viability, and gene expression assays (Aim 2) to determine exposure tolerance and molecular system activation. These efforts coordinate closely with advanced imaging studies (Aim 3) to visualize, and chemically and physically characterize, microbial reactions to nanoparticle exposure. There is close coordination between all aims in order to gain a clear understanding of how nanoparticle composition, size, and concentration affect microbial response. Correlations between microbial response and the physical and chemical properties of the nanoparticles are sought. This

**information will provide a foundation for understanding the potential fates of these materials in the environment, for guiding the development of effective nanoparticle-based technologies, and for understanding how microbial systems adapt to these exposures.**

Engineered nanostructures have a central role in energy conservation strategies and economic growth. One of the most significant impacts of engineered nanostructures is for effecting heterogeneous catalysis as required for fuel transformation, energy storage, polymer production and chemical synthesis. Metal and metal oxide nanoparticles are often used. They possess high surface areas and the ability to selectively mediate chemical transformations. The size and composition of the particle affects performance and may similarly affect nanoparticle fate and transport in the environment. The transformation of such nanoparticle catalysts in the environment is likely to proceed through interactions with bacteria. Several bacterial species are well known to interact with nanoparticles. Nanoparticle production, nanoparticle toxicity, nanoparticle binding and incorporation with bacteria have all been observed. However, basic knowledge that would allow prediction of the probable interaction between an engineered nanoparticle and bacteria is lacking. Our efforts seek to quantify and characterize interactions between engineered metal and metal oxide nanoparticles and selected microbial species. The effect of size and chemical composition of nanoparticles that are currently considered for various applied uses are being studied. Initial efforts are focused on the effects of cerium oxide nanoparticles on the growth, viability and genetic response of *E. coli*. Well-characterized CeO<sub>2</sub> nanoparticles of various sizes have been prepared and presented to bacterial cells in a dose dependent manner. Advanced imaging techniques are used to evaluate the binding and fate of the nanoparticles and the bacterial cell. The results of these studies will provide a basis for understanding how nanoparticle size and composition influence their interactions with microorganisms, and how microorganisms may alter the fate and transformation of engineered nanoparticles in the environment.