# Lawrence Berkeley National Laboratory

## 4

## VIMSS Computational Microbiology Core Research on Comparative and Functional Genomics

Adam Arkin*[1,2,3] (aparkin@lbl.gov), Eric Alm[1], Inna Dubchak[1], Mikhail Gelfand, Katherine Huang[1], Vijaya Natarajan[1], Morgan Price[1], and Yue Wang[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of California, Berkeley, CA; and [3]Howard Hughes Medical Institute, Chevy Chase, MD

**Background**. The VIMSS Computational Core group is tasked with data management, statistical analysis, and comparative and evolutionary genomics for the larger VIMSS effort. In the early years of this project, we focused on genome sequence analysis including development of an operon prediction algorithm which has been validated across a number of phylogenetically diverse species. Recently, the Computational Core group has expanded its efforts, integrating large amounts of functional genomic data from several species into its comparative genomic framework.

**Operon Prediction.** To understand how bacteria work from genome sequences, before considering experimental data, we developed methods for identifying groups of functionally related genes. Many bacterial genes are organized in linear groups called operons. The problem of identifying operons had been well studied in model organisms such as *E. coli*, but we wished to predict operons in less studied bacteria such as *D. vulgaris*, where data to train the prediction method is not available. We used comparisons across dozens of genomes to identify likely conserved operons, and used these conserved operons instead of training data. The predicted operons from this approach show good agreement with known operons in model organisms and with gene expression data from diverse bacteria.

**Statistical Modeling of Functional Genomics Experiments**. These operon predictions give hints to the function and regulation of many genes, but they can also aid the analysis of gene expression data. Genes in the same operon generally have similar expression patterns, so the degree to which genes in the same operon have correlated measurements gives an indication of the reliability of the data. Although most analyses of gene expression data have assumed that there are no systematic biases, we found that many data sets have systematic biases – biases that can not be corrected simply by increasing the number of experimental replicates. Using *a priori* knowledge of operon structure from our predictions, we can measure and account for these systematic biases, and more accurately assign confidence levels to experimental measurements. Furthermore, if several genes in an operon have consistent measurements, we have developed novel statistical models that assign much higher confidence to those measurements.

**Evolution of Microbial Genomes.** Our analysis of operons also led us to discoveries about how bacteria evolve. First, a popular theory has been that operons are assembled by horizontal gene transfer, and that operons exist, in part, to facilitate such transfers. We showed that such transfers are not involved in operon formation, and instead argue that operons evolve because they improve gene regulation. Second, we discovered that operons are preferentially found on the leading strand of DNA replication. (In most bacteria, a majority of genes are on the leading strand.) This observation



* Presenting author

is not explained the leading theories for strand bias. Instead, we note that genes, and especially long operons, are turned off during DNA replication, and these disruptions are shorter for operons on the leading strand. We believe that this mechanism can explain the known patterns of strand bias.

**Metabolic Reconstruction of Delta-Proteobacteria.** Species in the delta subgroup of the proteobacteria represent an important constituent of natural environmental diversity with key properties such as the ability to reduce heavy metals that make them of particular relevance to DOE core missions. Recently, a number of delta-proteobacterial genomes were sequenced, yet little is known about the physiology and regulation of key pathways. We have completed a comprehensive survey of regulatory signals and metabolic reconstruction of metal-reducing delta-proteobacterial species using comparative genomic analysis. In our survey, we characterized the evolution of 15 distinct regulons across six species. Interestingly, these species shared as many regulatory pathways in common with *B. subtilis*, a gram-positive bacterium, as they did with *E. coli*, itself a member of the proteobacteria. In addition to previously characterized regulons, we discovered a new CRP-like transcription factor that controls the sulfate-reduction machinery in *Desulfovibrio spp.*, and is generally present across anaerobic species, which we have named HcpR.

**Data Analysis.** The Computational Core group also played a role in the interpretation of experimental data generated by the Functional Genomics Core group. In a recent experiment in which *D. vulgaris* cells were subject to nitrite stress, the Computation Core group developed a detailed biological model that explains the observed transcriptional responses at a molecular level. In particular, enzymes involved in nitrite reduction to ammonia and incorporation of ammonia into glutamate were up-regulated, while the sulfate reduction machinery was down-regulated. In addition, iron uptake and oxidative stress genes were found to be up-regulated. Individual transcription factors along with their cognate DNA motifs were identified for each of these responses, and a model was proposed in which nitrite or other nitrogen intermediates play a role in oxidizing Fe(II), which in turn de-represses transcription from both the iron uptake and oxidative stress regulons.

**Data Management.** To support the larger VIMSS effort, the computational core group has deployed several new databases: the Biofiles database for rapid upload of arbitrary data types; the Experimental Data Staging and Experiment/Data Reporting Systems (EDSS/EDR) to automate the processing of key data types such as gene expression experiments; and the MicrobesOnline database which features a suite of analysis and visualization tools.

The EDSS database contains information and data from biomass production experiments (time points, stressor, direct cells counts, micrographs) and growth curve experiments. Several Web interfaces have been developed to access the EDSS database, including, details about the biomass production experiments (lab procedures, sample allocations, shipping conditions), tables of QA data (direct counts), and plots of growth curve data. In addition, time points and information about stressors stored in EDSS are accessed when the results of other experiments (e.g., microarray experiments) are analyzed and results compared. The EDR database and Web interface were developed to provide a reporting system to track data generation from the starting point of biomass production through the entire suite of laboratory analyses performed on the biomass. The reporting system allows PIs to document each step in the experimental pipeline (e.g., sample preparation, QA measurements, etc.). A major component of the EDR system is a Web interface for writing and submitting reports about data being uploaded to the VIMSS file server. The interface requires users to describe the laboratory analysis that generated the data (type of analysis, dates data were generated, biomass source, etc.), content of the uploaded data, the file format and the format of the data within the file(s), and any reference information needed to fully understand the data file(s).

**The MicrobesOnline Database.** The MicrobesOnline database currently hosts 180 genomes and features a full suite of software tools for browsing and comparing microbial genomes. Highlights include operon and regulon predictions, a multi-species genome browser, a multi-species Gene Ontology browser, a comparative KEGG metabolic pathway viewer and the VIMSS Bioinformatics Workbench for more in-depth sequence analysis. In addition, we provide an interface for genome annotation, which like all of the tools reported here, is freely available to the scientific community. To keep up with the ever-increasing rate at which microbial genomes are being sequenced, we have established an automated genome import pipeline. Since August 2004 this automated pipeline has allowed us to increase the number of hosted genomes from 100 to 180.

A number of outside groups are currently using the MicrobesOnline database for genome annotation projects. To facilitate the use of this community resource we are developing a sophisticated access control system, so individual research groups can use the power of the VIMSS annotation tools, while keeping data from their own particular genome project private until their analyses are ready to be made public.

**Addition of Functional Genomics to MicrobesOnline.** In addition to browsing comparative genomics, the MicrobesOnline database and website now allows users to browse and compare functional genomics data. In particular we have started with gene expression microarray data as a test case for high-throughput functional genomics measurements. Currently gene expression data from 262 experiments across four different species are hosted in the database. Software tools available from the MicrobesOnline functional genomics web portal allow users to overlay expression data on predicted operon structure or metabolic pathways. In addition, an operon-based estimate of microarray accuracy has proven useful in determining the quality of experimental measurements.

# 5

## The Virtual Institute of Microbial Stress and Survival (VIMSS): Deduction of Stress Response Pathways in Metal/Radionuclide Reducing Microbes

Carl Abulencia[4], Eric Alm[1], Gary Andersen[1], Adam Arkin[1]* (APArkin@lbl.gov), Kelly Bender[5], Sharon Borglin[1], Eoin Brodie[1], Swapnil Chhabra[3], Steve van Dien[6], Inna Dubchak[1], Matthew Fields[7], Sara Gaucher[3], Jil Geller[1], Masood Hadi[3], Terry Hazen[1], Qiang He[2], Zhili He[2], Hoi-Ying Holman[1], Katherine Huang[1], Rick Huang[1], Janet Jacobsen[1], Dominique Joyner[1], Jay Keasling[1], Keith Keller[1], Martin Keller[4], Aindrila Mukhopadhyay[1], Morgan Price[1], Joseph A. Ringbauer, Jr.[5], Anup Singh[3], David Stahl[6], Sergey Stolyar[6], Jun Sun[4], Dorothea Thompson[2], Christopher Walker[6], Judy Wall[5], Jing Wei[4], Denise Wolf[1], Denise Wyborski[4], Huei-che Yen[5], Grant Zane[5], Jizhong Zhou[2], and Beto Zuniga[6]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]Sandia National Laboratories, Livermore, CA; [4]Diversa, Inc., San Diego, CA; [5]University of Missouri, Columbia, MO; [6]University of Washington, Seattle, WA; and [7]Miami University, Oxford, OH

### Introduction

The mission of the Virtual Institute of Microbial Stress and Survival, is to understand the molecular basis for the survival and growth of microbes in the environment. Towards this end VIMSS has



* Presenting author

designed a series of key protocols, experimental pipelines and computational analysis to support and coordinate research in this area. Our flagship project aims to elucidate the pathways and community interactions which underlie the ability of *Desulfovibrio vulgaris* Hildenborough (DvH) to survive in diverse, possibly contaminated environments and reduce metals. Their ability to reduce toxic Uranium and Chromium, major contaminants of industrial and DOE waste sites, to a less soluble form has made them attractive from the perspective of bioremediation.

We are discovering the molecular basis for the physiology of these organisms first through characterization of the biogeochemical environment in which these microbes live and how different features of these environments affect their growth and reductive potential. We have created an integrated program through the creation of an experimental pipeline for the physiological and functional genomic characterization of microbes under diverse perturbations. This pipeline produced controlled biomass for a plethora of analyses as described below and is managed through workflow tools and a data management and analysis system. The effort is broken into three interacting core activities: The Applied Environmental Microbiology Core; the Functional Genomics Core; and the Computational Core.

**Accomplishments of the Applied Environmental Microbiology Core (AEMC)**

**Characterization of the Environment.** The AEMC has collected or completed basic analysis of the stressors present at a number of NABIR FRC site, and characterized the microbial community before and after stimulation using 16SRNA microarrays. Large insert cloning was used to characterize the enrichment of genomic functions in these environments. Diversity analysis of library clones revealed genes used in transport, small molecule binding, toxicity response and DNA synthesis, among others. We are now targeting primers for enrichment of signal transduction pathway components. In addition, nine *D. vulgaris*-like bacteria (DP1-9) were isolated from a metal impacted field site (Lake DePue, Illinois). All had identical 16S rRNA and *dsrAB* genes that were virtually identical to the orthologous genes of DvH. Complementary whole-genome microarray hybridization revealed that approximately 300 deleted genes were distributed in six regions of the chromosome, annotated as conserved/ hypothetical or phage related genes in DvH. We are now following up characterization of these phageless strains.

**Biomass Production and Characterization:** In the core pipeline experiments each microbe is first characterized physiologically using Omnilog phenotypic microarrays. A stressor condition is then applied to a large set of batch cultures and samples are collected periodically to obtain a time-series of cellular response. Each time-point is split so that the cells can be imaged, analyzed through synchrotron IR microscopy to measure the bulk physiological changes of the cells during their response, and determine the optimal time points to send to the functional genomics core (FGC) for transcript, protein and metabolite analysis. Response to oxygen stress, salt stress (shock and adaptation) and nitrate have been fully characterized in this way. In related work, we are developing laboratory systems that simulate environmental conditions than can not be achieved in pure culture, initially focusing on co-cultures of two different *Desulfovibrio* species (*DvH* and *Desulfovibrio sp*. PT2) syntrophically coupled to a hydrogenotrophic methanogen (*Methanococcus maripaludis*). Transcriptional dynamics of the co-culture has been measured by the FGC. In addition, a metabolic stoichiometric model has been constructed using flux balance analysis (FBA) to complement and direct experimental studies on the physiology of *DvH* growing either alone or in co-culture.

**Accomplishments of the Functional Genomics Core**

**Genetics:** To improve the genetic accessibility of *DvH*, we found the cells to be sensitive to the antibiotic Geneticin or G418, therefore, allowing kanamycin resistance to be used as a genetic marker.

Using the modified mini-Tn5 from Bill Metcalf, we have been able to generate a library of transposon mutants that appear to be randomly inserted throughout the genome. Several putative regulatory genes were among those mutated and we are screening for mutants of specific phenotypes. We have generated tagged hspC and rpoB genes in single copy controlled by their native promoters to use for development of assays for protein complexes. We have established a procedure for making gene deletions in non-essential genes that introduces a unique oligonucleotide that can be used for mutant identification. With this procedure, we have generated a putative *fur* deletion that is increased four fold over the wild type in its resistance to manganese. We are also generating a library of histidine kinase (HK) knockouts. *DvH* has 69 HKs that govern signal transduction. A suicide vector has been designed and created to enable gene deletion and concurrent "bar-coding" of the chromosome. Our preliminary results include 6 potential knock-out mutants.

Transcriptomics: We have, to date, characterized five stresses in *DvH* and five in *S. oneidensis* and results are integrated with the VIMSS MicrobesOnline Database. New regulons and their cis-regulatory sequences have been discovered along with new hypotheses of the pathways by which both organisms respond to these different stressors. A number of papers are in press, submitted or are in preparation around this topic.

**Proteomics:** We have developed three complementary proteomics methods to characterize protein expression in our microbes Differential In Gel Electrophoresis followed by MALDI-TOF and nanLC-ISI-QTOF, Isotope coded affinity tagging with tandem LC mass spec, and direct MS-MS. In addition, to characterize protein complexes we have developed both a high throughput cloning & expression of DvH proteins in *E. coli* and methods for expression of genetically-modified proteins at their native levels in the host organism. These proteins are then used as bait proteins to enable "pull-down" of associated proteins.

**Metabolomics:** We have set up and optimized both Capillary electrophoresis (CE) and Liquid chromatography (LC) coupled with Mass spectrometry (MS) methods for characterization of metabolites. Metabolite extraction protocols have been developed for *DvH*.

**Accomplishments of the Computational Core**

During the past year the computational core has focused on building the comparative and functional genomic analysis tools to aid in the prediction of regulatory networks in microbes, elucidate their evolutionary relationships and extract the most meaning from the functional genomics and phenotypic data described in the last two sections. We have developed an increasingly sophisticated experimental and data management system that centralizes and serves all VIMSS data and tracks the progress through experimental runs of the pipeline. One of the key technologies we have developed is a set of web-accessible comparative genomic tools (http://vimss.org) designed to facilitate multi-species comparison among prokaryotes. Highlights of the system accessible through the VIMSS website include operon and regulon predictions based on novel methods we have proven to work on a wide diversity of micro-organisms, a multi-species genome browser, a multi-species Gene Ontology browser, a comparative KEGG metabolic pathway viewer and the VIMSS Bioinformatics Workbench for in-depth sequence analysis. In addition, we provide an interface for genome annotation, which like all of the tools reported here, is freely available to the scientific community. This tool has been used successfully by a number of projects. In particular, an Joint Genome Institute Annotation Jamboree we ran to annotate *D. desulfuricans* G20 which will likely be reclassified as *D. alaskensis*. We have also been working on tools for modeling pathways and understanding how the molecular strategies we measure in the lab confer the ability to survive in the environment.



* Presenting author

# 6

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

Terry C. Hazen*[1] (tchazen@lbl.gov), Carl Abulencia[3], Gary Andersen[1], Sharon Borglin[1], Eoin Brodie[1], Steve van Dien[5], Matthew Fields[6], Jil Geller[1], Hoi-Ying Holman[1], Rick Huang[1], Janet Jacobsen[1], Dominique Joyner[1], Martin Keller[3], Aindrila Mukhopadhyay[1], David Stahl[5], Sergey Stolyar[5], Jun Sun[3], Dorothea Thompson[2], Judy Wall[4], Denise Wyborski[3], Huei-che Yen[4], Grant Zane[4], Jizhong Zhou[2], and Beto Zuniga[5]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]Diversa, Inc., San Diego, CA; [4]University of Missouri, Columbia, MO; [5]University of Washington, Seattle, WA; and [6]Miami University, Oxford, OH

**Field Studies**

*Identification of Different Relationships Between Contaminated Groundwater Samples Based Upon Geochemical Data or Multiple Gene Sequences from Microbial Communities.* Factor analysis was used to identify a subset of variables that may explain a majority of the observed variance between the contaminated groundwater sites, and principal components analyses were then used to compare the sites based upon geochemistry, phylogenetic markers (n=353), and functional markers (n=432). The clonal libraries of the multiple genes (SSU rRNA gene, *nir*K, *nir*S, *amo*A, *pmo*A, and *dsr*AB) were constructed from groundwater samples (n=6) that varied in degrees of contamination. When geochemical characteristics were analyzed, the data suggested that the samples could be differentiated based upon pH, nitrate, sulfate, nickel, aluminum, and uranium. Similar relationships between the sites were observed when 107 analytes were used, but more resolution was achieved between the more contaminated sites. In addition, a majority of the variance between the acidic samples could be accounted for by tetrachloroethene, $^{99}Tc$, $SO_4$, Al, Th and 1,1,2-trichloro-1,2,2-trifluoroethane. The analysis based on a phylogenetic marker resulted in different groupings for background and the two circumneutral sites compared to the geochemical analysis, and analyses of the OTU distributions for the functional genes each predicted different relationships between the sites. A tripartite PCA explained 76% of the variance and grouped the background sample with the three, heavily contaminated sites. When all gene OTUs were used in the analyses, the sites were more similar than in any other comparison, 94% of the observed variance cold be explained, the background site was grouped with the contaminated sites, and possible key populations were identified by factor analysis. The data suggested that even though the background site was phylogenetically and geochemically distinct from the acidic sites, the extreme conditions of the acidic samples might be more analogous to the limited-nutrient conditions of the background site.

*Biopanning/Clone libraries*. Diversa extracted high molecular weight DNA from organisms present in contaminated soil sediment samples using a method that preserves the integrity of the DNA. Because the number of organisms in these samples was low, the genomic DNA was amplified using a phage polymerase amplification system. 16S rRNA analysis was then used to examine the microbial diversity of the samples. The amplified DNA was also used in the construction of large and small insert DNA libraries. These libraries were then screened for the presence of histidine kinase genes with homology to a subfamily of *Desulfovibrio vulgaris* histidine kinases. Genomic DNA has been extracted and amplified from nine different sites at the NABIR field research center. 16S rRNA analysis revealed the presence of distinct bacterial phyla, including proteobacteria, acidobacteria,

and planctomycetes. Small and large insert libraries were constructed for all samples and examined for clonal diversity. Plaque hybridization of these libraries to histidine kinase homologous probes resulted in multiple positive clones. These clones will be compared and used to develop a better understanding of cellular responses to different environmental factors. These experiments have furthered the understanding of how the biological organisms in a contaminated system are organized, regulated and linked.

*Enrichments.* Nine *D. vulgaris*-like bacteria (DP1-9) were isolated from a metal impacted field site (Lake DePue, Illinois) as an additional reference set for comparative stress analyses. All had identical 16S rRNA and *dsrAB* genes that were virtually identical to the orthologous genes of *D. vulgaris* Hildenborough (DvH). However, pulse field gel electrophoretic analysis of I-CeuI digests identified a large deletion in the genomes of all isolates. Complementary whole-genome microarray hybridization revealed that approximately 300 deleted genes were distributed in six regions of the chromosome, annotated as conserved/ hypothetical or phage related genes in DvH. These deletions were also confirmed by PCR analysis, using primers complementary to regions flanking the deletions. Continuing collaboration with Judy Wall (U Missouri) has shown that one of the "phage-deficient" *D. vulgaris* strains (DP4) serves as host for latent viruses of *D. vulgaris* Hildenborough, identifying two phage morphotypes by EM. MPN enrichments from FRC area 2 sediments were developed using a PIPES buffered B2 medium supplemented with: 1) lactate, 2) lactate plus ethanol, 3) acetate, 4) propionate 5) pyruvate or 6) hydrogen plus carbon dioxide. All showed sulfate reducing activity within a range of $10^{-1}$ to $10^{-4}$ dilutions. Thirty isolates from the lactate medium were shown by 16S rRNA sequence to be affiliated with the "*Firmicutes*". A Gram-negative sulfate reducer (curved-rod morphology) maintained on an $H_2/CO_2$ plus acetate medium was also isolated.

*Dual culture systems*. The kinetics and stoichiometry of syntrophic growth were determined in batch culture by quantifying each population, substrate consumption (lactate), evolution of metabolic intermediates ($H_2$ and acetate), and end-product accumulation ($CO_2$ and methane). *D. vulgaris* monocultures were grown at generation times comparable to syntrophic batch cultures (24 and 36 hours) in sulfate-limited chemostats for comparative transcription analyses. Fermentative growth *D. vulgaris* on a lactate medium (sulfate minus) with continuous headspace purging was also developed for comparison. Transcription analyses of co-cultures identified a preliminary set of *D. vulgaris* genes either up or down regulated with syntrophic association, including periplasmic and cytoplasmic hydrogenases. These analyses are now being replicated at ORNL. A metabolic stoichiometric model was constructed using flux balance analysis (FBA) to complement and direct experimental studies on the physiology of *D. vulgaris* growing either alone or in co-culture. The network for each organism was based primarily on the annotated genome sequences, supplemented by available biochemical knowledge. The *Desulfovibrio* model consists of 86 reactions and 73 internal metabolites, while that of the methanogen contains 84 reactions and 72 metabolites.

**Stress Experiments**

*High Throughput Biomass Production*. Producing large quantities of high quality and defensibly reproducible cells that have been exposed to specific environmental stressors is critical to high throughput and concomitant analyses using transcriptomics, proteomics, metabolomics, and lipidomics. Culture of *D. vulgaris* is made even more difficult because it is an obligate anaerobe and sulfate reducer. For the past two years, our Genomics:GTL VIMSS project has developed defined media, stock culture handling, scale-up protocols, bioreactors, and cell harvesting protocols to maximize throughput for simultaneous sampling for lipidomics, transcriptomics, proteomics, and metabolomics. All cells for every experiment, for every analysis are within two subcultures of the original ATCC culture of *D.*



* Presenting author

*vulgaris*. In the past two years we have produced biomass for 38 integrated experiments (oxygen, NaCl, $NO_3$, $NO_2$, heat shock, cold shock, pH) each with as much as 30 liters of mid-log phase cells ($3 \times 10^8$ cells/ml). In addition, more than 40 adhoc experiments for supportive studies have been done each with 1-6 liters of culture. All cultures, all media components, all protocols, all analyses, all instruments, and all shipping records are completely documented using QA/QC level 1 for every experiment and made available to all investigators on the VIMSS Biofiles database (http://vimss.lbl.gov/perl/biofiles). To determine the optimal growth conditions and determine the minimum inhibitory concentration (MIC) of different stressors we adapted plate reader technology using Biolog and Omnilog readers using anaerobic bags and sealed plates. Since each well of the 96-well plate produces an automated growth curve, over more than 200 h, this has enabled us to do more than 4,000 growth curves over the last two years. Since the Omnilog can monitor 50 plates at a time, this allows us to do more than 5,000 growth curves in a year. We have also developed chemostat techniques using a specially made extremophile fermentor (FairManTech) that has no internal metal parts. With this system we can get *D. vulgaris* to steady state from the freezer in less than 80 h in turbidostat mode, with a dilution rate of 0.25 l/h. Each reactor has a useable volume of 3 liters, with our current two reactors this enables production of 6 liters of steady state culture twice a week. We have also developed new harvesting techniques to minimize the stress caused by sample preparation for shipping. Since the volumes being centrifuged are large, the cells were not cooling fast enough to ensure high quality samples, so we devised a sampling apparatus that draws the cells from the culture vessel through capillary tubing in a MgCl ice bath that lowers the sample to 4°C in less than 20 sec. These procedures have maximized our reproducibility and throughput for the 8 labs involved.

*Phenotypic Responses.* Phenotypic Microarray™ analysis is a recently developed analytical tool to determine the phenotype of an organism. The plates, which are commercially available from Biolog™ (Hayward, CA), consist of 20 96-well plates. The first eight plates test a variety of metabolic agents, including electron donors, acceptors, and amino acids. Plates 9 and 10 cover a pH and osmotic stressors, while plates 11-20 contain a variety of inhibitors, including toxic agents and antibiotics. We have developed the ability use these plates under anaerobic conditions by inoculating plates in an anaerobic chamber, and heat-sealing them in polyethylene bags containing an anaerobic sachet. Using this technique, anaerobic conditions were maintained for up to a week. It was found that preconditioning of the cells in specialized media was required for the different types of plates in order to get a valid phenotype. The plates have been successfully used to characterize the phenotype of the *D. vulgaris* ATCC strain and are currently being applied to mutant strains to provide rapid screening of mutant phenotypic changes, for rapid pathway analyses and modeling. See (https://vimss.lbl.gov/~jsjacobsen/cgi-bin/Test/HazenLab/Omnilog/home.cgi) for sample data sets and analyses.

*Synchrotron FTIR Spectromicroscopy for Real-Time Stress Analysis*. LBNL's newly developed synchrotron radiation-based (SR) Fourier-transform infrared (FTIR) spectromicroscopy beamline allows the study of many biochemical and biophysical phenomena non-invasively as the processes are happening. It has been enabling for our high throughput determinations of optimal sampling times stress experiments. We can observe real-time changes of major biomolecule pools within cells as they are exposed to different stressors. This allows us to pick optimal sampling points during the stress response for transcriptome, protemome, meatabolome, and lipidome analyses. It also allows us to verify the purity and state of the culture for QA/QC.

7

# VIMSS Functional Genomics Core: Analysis of Stress Response Pathways in Metal-Reducing Bacteria

Aindrila Mukhopadhyay[1], Steven Brown[4], Swapnil Chhabra[2], Brett Emo[3], Weimin Gao[4], Sara Gaucher[2], Masood Hadi[2], Qiang He[4], Zhili He[4], Ting Li[4], Yongqing Liu[4], Alyssa Redding[1], Joseph Ringbauer, Jr.[3], Dawn Stanek[4], Jun Sun[5], Lianhong Sun[1], Jing Wei[5], Liyou Wu[4], Huei-Che Yen[3], Wen Yu[5], Grant Zane[3], Matthew Fields[4], Martin Keller[5] (mkeller@diversa.com), Anup Singh[2] (aksingh@sandia.gov), Dorothea Thompson[4], Judy Wall[3] (wallj@missouri.edu), Jizhong Zhou[4] (zhouj@ornl.gov), and Jay Keasling[1]* (keasling@socrates.berkeley.edu)

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Sandia National Laboratories, Livermore, CA; [3]University of Missouri, Columbia, MO; [4]Oak Ridge National Laboratory, Oak Ridge, TN; and [5]Diversa Inc., San Diego, CA

The Functional Genomics Core is part of the VIMSS project and there is a separate overview presentation of the VIMSS project. Environmental contamination by metals and radionuclides constitutes a serious problem in many ecosystems. Bioremediation schemes involving dissimilatory metal ion-reducing bacteria are attractive for their cost-effectiveness and limited physical detriment and disturbance on the environment. *Desulfovibrio vulgaris, Shewanella oneidensis*, and *Geobacter metallireducens* represent three different groups of organisms capable of metal and radionuclide reduction whose complete genome sequences were determined under the support of DOE-funded projects. Utilizing the available genome sequence information, we have focused our efforts on the experimental analysis of various stress response pathways in *D. vulgaris* Hildenborough using a repertoire of functional genomic tools and mutational analysis.

Originally isolated in 1946, from the clay soils in Hildenborough, Kent (UK), *Desulfovibrio vulgaris* Hildenborough belongs to a class of sulfate reducing bacteria (SRB) that are found ubiquitously in nature. As with most soil bacteria that do not live permanently in hyperosmotic environments, a NaCl salt stress in a *D .vulgaris* can be expected to result in at least two primary responses, osmotic and that towards Na+ ions. The genomic sequence of *D. vulgaris* indicates that a variety of mechanisms may be employed to counter these two stresses. In order to understand these mechanisms at the physiological level an integrated functional genomics analysis was conducted. Data from microarray analysis of the transcriptome, quantitative and qualitative proteomic analysis, PLFA profiling and IR studies reveal many interesting responses to stress. For example, with respect to hyperosmotic stress, the three gene operon that regulates the uptake of the osmoprotectant, glycine betaine is highly up-regulated. With respect to Na+ stress, Na+/H+ antiporters such as the dehydrogenase *mnhA* show upregulation in mRNA levels. As might be expected with cellular physiology, a myriad of other relevant responses were observed such as upregulation in ATP synthesis, down-regulation in flagellar systems. IR studies also indicate changes in cell wall composition. Moreover, several genes of unknown function were observed to be significantly and reproducibly changing, and may lead to the annotation of additional candidates involved in Salt stress. The study also attempts to understand the general correlation of proteomics vs. transcriptomics data. Results from this work will lead to further studies with metabolic profiling, and gene deletion mutants.



* Presenting author