Environmental Genomics

B1

Identification and Isolation of Active, Non-Cultured Bacteria from Radionuclide and Metal Contaminated Environments for Genome Analysis

Cheryl R. Kuske[1] (kuske@lanl.gov), Susan M. Barns[1], and Leslie E. Sommerville[1,2]

[1]Los Alamos National Laboratory; and [2]Ft. Lewis College, Durango, Colorado

The overall goals of this project are to identify active, non-cultured soil bacteria present in radionuclide and metal contaminated environments, and to collect bacterial cells representing abundant, non-cultured populations for genome analysis. Our studies have focused on non-cultured members of the *Acidobacterium* division. Members of this division are widespread in contaminated and pristine soils having vastly different physical and chemical characteristics, and they have been found to represent a major fraction of the non-cultured bacteria in several soils (by 16S rRNA clone library analysis). Toward accomplishing the above goals we have made progress on five specific objectives.

Objective 1: Analysis of soil bacterial community rRNA. Historically, rRNA gene-based surveys of bacteria have been conducted on the pool of DNA encoding the rRNA gene. This provides a snapshot of the total composition of a sample, but does not indicate which members may be active at the time of sampling. We have conducted parallel DNA- and RNA-based analyses from a surface soil and a bacterial cell preparation extracted from that soil. By comparing the DNA-based composition to a parallel rRNA-based analysis, we hope to identify which components of the total community are active (ie. contain abundant ribosomes).

Objective 2: Survey of *Acidobacterium* division members in contaminated soils. Our second objective is to conduct a survey of DOE sites contaminated with radionuclides and metals to determine the diversity of *Acidobacterium* division bacteria and identify active members of this division in contaminated soils. Collections are in progress from the NABIR FRC (Oak Ridge, TN), PNNL, an UMPTRA site in Rifle, CO, and the Nevada Test Site. Through, we are assessing the presence and diversity of *Acidobacterium* division members using simultaneous DNA and RNA extraction, followed by 16S rRNA PCR and RT-PCR for rDNA and rRNA, respectively. Analysis of contaminated and background sites from the FRC indicate (a) subsurface biomass was very low and only DNA was recoverable in sufficient quantities for analysis from this site. (b) The composition of *Acidobacterium* division bacteria is similar across replicates of the background site, but is very different from the contaminated sites (saturated zone near wells TPB15, PTB16, DB13 in Area 2). (c) The contaminated sites contain very diverse *Acidobacterium* division members. Two new subgroups comprise the majority of 16S rRNA clones from the sample taken near well TPB16. A significant number of clones for one of the subgroups were also present in the other two contaminated sites. Clones for these new subgroups have not been found in samples from the background site.We are continuing to collect surface and subsurface samples from the other field sites for RNA- and DNA-based composition analyses.

Objective 3: Collection of *Acidobacterium* division members from soil for genome analysis. We have continued efforts to develop sensitive, specific hybridization methods for detection of *Acidobacterium* division members in environmental samples using riboprobes, and to collect hybridized cells from soil bacteria using flow cytometry cell sorting. We have developed a series of specific hybridization probes for the division and some of its major subgroups. In collaboration with Hong Cai (LANL), we determined that although cells could be specifically labeled and observed microscopically, the fluorescence intensity of each cell was too low to allow cell sorting on the LANL instrument. We are currently working with Diversa Corp. and PNNL to conduct cell sorting experiments on a Diversa instrument with better calibration for bacterial cells. At LANL, we are using gradient centrifugation to obtain

Acidobacterium division-enriched bacterial cell preparations for mixed genome libraries.

Objective 4: Culture of *Acidobacterium* division species from soils. In collaboration with Martin Keller (Diversa Corp.) and Fred Brockman (PNNL) we are using Diversa's Gel Microdroplet (GMD) technology in attempts to culture *Acidobacterium* division cells from our soil samples. Diversa has attempted culture in a matrix of different media and aeration conditions. To date we have several candidate cultures representing subgroups 1 and 6. DNA extracted from cultured cells will be used to generate genomic libraries at LANL.

Objective 5: Analysis of contaminated soil microcosm RNAs. We will use ribosomal RNA (to determine species composition) and messenger RNA (to determine active functions) to examine bacterial community response to radionuclide contamination in soil microcosm experiments. In collaboration with Mary Neu (LANL), we are setting up preliminary soil microcosms to support method development and for preliminary analysis of functional response of the natural soil bacterial community to different forms of Pu and U.

B3

A Metagenomic Library of Bacterial DNA Isolated from the Delaware River

David L. Kirchman, (kirchman@udel.edu), Matthew T. Cottrell, and Lisa Waidner

College of Marine Studies, University of Delaware, Lewes, DE 19958

Most bacteria and archaea in natural environments still cannot be isolated and cultivated as pure cultures in the laboratory, and the microbes that can be cultured appear to be quite different from uncultured ones. Consequently, the phylogenetic composition, physiological capacity and genetic properties of natural microbes have to be deduced from fluorescence *in situ* hybridization (FISH) assays and bulk properties of microbial assemblages, and from a variety of PCR-based methods applied to DNA isolated directly from natural samples. Another culture-independent approach is to clone this DNA directly into appropriate vectors and to screen the resulting "metagenomic library", which theoretically consists of all possible genes from the microbial assemblage. We applied this general approach to a sample from the freshwater end of the Delaware Estuary as part of our efforts to understand carbon and nitrogen cycling in environments like estuaries with large environmental gradients. Metagenomic libraries have been constructed for soils and marine samples, but not for freshwaters. High molecular weight DNA from the bacterial size fraction was isolated and cloned into the fosmid vector pEpiFOS-5 (Epicentre). Our library consisted of 4608 clones with an average insert size of 40 kB, representing about 90 genomes, if we assume a genome size of 2 mB.

Screening the library revealed several surprises, including genes found previously in metagenomic libraries of oceanic samples. Our library appears to be dominated by *Cytophaga*-like bacteria according to the 16S rRNA data collected by DGGE analysis of PCR amplified 16S rRNA genes. Of the 80 clones bearing 16S rRNA genes, about 50% appear to be from the *Cytophaga-Flavobacteria*, a complex cluster in the Bacteroidetes division. FISH analysis of the original microbial assemblage indicated that *Cytophaga*-like bacteria were only about 15% of the community. The next most abundant 16S rRNA genes in the library are from G+ *Actinobacteria*, which others have shown to be abundant in freshwater lakes. But beta-proteobacteria usually dominate freshwater systems and were the most abundant group in our sample according to the FISH analysis, yet beta-proteobacteria accounted for only about 15% of the 16S rRNA genes in the metagenomic library, much less than the 25% found by FISH.

The library was also screened for hydrolysis of the fluorescent analog of cellulose, MUF-beta-1,4-glucoside. Twenty-four of the 2,400 clones screened had cellulase activity, which was inferred from rates of analog hydrolysis 2.5-fold greater than the control with vector alone. The activity of seven clones was 3-fold greater than the control, while 40 additional clones had activities between 2 and 2.5-fold higher than the control. The variation of activities observed suggests that the library contains genes encoding variety of glycosyl hydrolases capable of cleaving this fluorgenic analogue. One of the cellulase-active clones was determined to harbor a 16S rRNA gene from a *Cytophaga*-like bacterium. This clone is now being completely sequenced by John Heidelburg (TIGR).

We also screened the library for genes indicative of two newly-discovered photoheterotrophic

metabolisms. Our fosmid library does not appear to contain the proteorhodopsin gene, which had been found by Beja et al. (Science (1999) 289: 1902-1906) in a marine metagenomic library. However, we found two clones that contain *pufM* and *pufL*, which code for reaction center proteins in bacteria carrying out anoxygenic photosynthesis. Although well known to be present in anoxic environments, the biophysical evidence of Kolber et al. (Science (2001) 292: 2492-2495) indicates that aerobic anoxygenic photosynthesizing bacteria are also present and perhaps are biogeochemically important in the oxic habitats, such as the oceans. One 33 kb clone from our library has a *pufL* sequence most similar to the protein sequence of the gamma-proteobacterium *Allochromatium vinosum*, whereas *pufM* of this clone is most similar to the freshwater beta-proteobacterium *Rhodoferax fermentans*. The second fosmid clone contains a 35 kb insert with the *Puf*L-M reaction center complex of alpha-proteobacterial origin. The *Puf*L sequence is most similar to that of alpha-4 proteobacteria subgroup, whereas the *Puf*M protein sequence is related to Monterey Bay environmental BAC clones and Monterey Bay isolates in the alpha-proteobacterial *Roseobacter* clade.

The diversity of rRNA genes, enzyme activities and *puf* genes in this freshwater library is consistent with our expectation that freshwater environments harbor diverse assemblages of microbes. The large number of clones with *Cytophaga*-like 16S rRNA sequences and with apparent glycosyl hydrolase activity is encouraging. Additional screening of hydrolase-active clones by fosmid-end sequencing may provide additional links to clones representing *Cytophaga*-like bacteria, which would support our efforts to explore the hydrolytic capabilities and DOM cycling by this important group of aquatic bacteria.

B5

Approaches for Obtaining Genome Sequence from Contaminated Sediments Beneath a Leaking High-Level Radioactive Waste Tank

Fred Brockman[1] (fred.brockman@pnl.gov), Margaret Romine[1], Kristin Kadner[2], Paul Richardson[2], Karsten Zengler[3], Martin Keller[3], and Cheryl Kuske[4]

[1]Pacific Northwest National Laboratory; [2]DOE Joint Genome Institute; [3]Diversa Corporation; and [4]Los Alamos National Laboratory

The SX Tank Farm at the U.S. Department of Energy's Hanford Site was built in 1953 to receive high level radioactive waste, and consists of one million gallon enclosed tanks. The waste resulted from recovery of purified plutonium and uranium from irradiated production fuels using methyl isobutyl ketone, aluminum nitrate, nitric acid, and sodium dichromate. Between 1962 and 1969, tens of thousands of gallons of radioactive liquid leaked from tank SX-108. An extreme environment formed in the vadose zone from incursion of radioactive, caustic, and toxic contaminants and heating from the self-boiling contents of the tank. Samples contain up to 50 microCuries of Cesium-137 per gram sediment, nitrate at 1% to 5% of sediment mass, pH's to 9.8, and were heated to 50 to 70 degrees C. These samples are the most radioactive sediments studied at the DOE Hanford Site in Washington state, and we hypothesized selection would result in a relatively nondiverse community. Previous work demonstrated the presence of a low number of cultured organisms from these sediments (Fredrickson et al, to be submitted).

The vast majority of microbial diversity in environmental samples has proved refractory to cultivation and therefore genome, proteome, and metabol- omic analysis. New strategies are needed to access the repertoire of genes, proteins and metabolic capabilities embodied in the community's genomic sequences. The project goal is to demonstrate an approach to obtain genetically-linked genome sequence from members of this low-biomass community. Our initial approach was to screen sediments and enrichments for the presence of communities *dominated* by a very few microbes representing *uncultured or poorly cultured* divisions (Hugenholtz et al, 1998). The purpose of the screening was to identify an appropriate

sample for constructing a community BAC library and then performing high throughput sequencing of BAC ends to assemble >1,000 Kbp of genetically linked sequence.

To evaluate the utility of this approach we first characterized the environmental DNA from 8 sediment samples and 91 enrichments by 16S amplification with conserved primers followed by sequencing of clones at the DOE Production Genomics Facility. We also performed PCR's on the samples using primers targeting 13 divisions of uncultured or poorly cultured bacteria, and cloned and sequenced successful PCR's. The results showed (1) only one sequence out of over 9,000 clones showed a BLAST hit to uncultured or poorly cultured bacteria, (2) *Archaea* 16S sequences could not be amplified, (3) biomass was about 10^5 viable cells per gram sediment, and (4) the depth of community penetration in the sediments was poor because the detection level (80,000 copies per gram sediment) was only 2 to 3 times higher than the indigenous template concentrations (determined by competitive PCR). Never the less, sequencing showed between 3 and 15 putative genera per sediment, and two of the highly contaminated sediments were dominated by alpha and/or beta proteobacteria. Proteobacteria are not primary inhabitants of Hanford Site deep vadose zone sediments, suggesting *in situ* microbial growth on components of the waste.

Because the results showed that our initial approach for studying genomes of uncultured and poorly cultured microbes at this site was not feasible, we pursued a novel culturing approach developed by Diversa Corporation. SX-108 sediments were grouped into high rad-low nitrate; high rad-high nitrate; low rad-high nitrate; and low rad-low nitrate groups. Nearby sediments recovered from similar depths were pooled to provide an uncontaminated control sample. As a positive control, a soil sample from Los Alamos Nat'l Lab. with a natural population known to be comprised of >25% Acidobacterium cells was prepared in collaboration with Cheryl Kuske. Cells were purified and concentrated from each sample using multiple nycodenz gradient centrifugations. Single cells were encapsulated in individual gel microdroplets (gmd's) and the community reconstituted by placing gel microdroplets into a column. The community was grown under low nutrient flux conditions and gmd's sorted by flow cytometry using intrinsic forward and side scatter to detect those containing microcolonies of, on average, 50 to 200 cells. Key aspects of this technology are that it enables propagation of single organisms with extremely slow growth rates, and preserves some of the community interactions and other specific requirements needed for successful cultivation.

For the positive control sample, sorted gmd's were screened with primers specific for Acidobacterium division. 16S sequences were amplified from positive gmd's and a number of the sequences group in a phylogenetic tree with subgroup 6 of the Acidobacterium division (see poster by Cheryl Kuske and Sue Barnes). This represents the first known culturing of these bacteria. Our next steps are to extract DNA by either standard techniques or employing whole genome amplification from small numbers of cells, and perform partial genome sequencing.

A total of 14,000 gmd's with putative microcolonies have been sorted from the four sets of pooled SX-108 sediments and from the uncontaminated control. In the previous study, less than 50 isolates were obtained from plating these same samples. A subset of these gmd's will be amplified, cloned, and sequenced to compare the community structure in these samples to one another, to the previously obtained isolates, and to the 16S sequences obtained by direct extraction of DNA from the sediments. Our hypothesis is that the gmd microcolonies will represent microbes from a number of poorly cultured or uncultured divisions. One or more of the most unique microbes will be used for partial genome sequencing as outlined above.

B7

Ecological and Evolutionary Analyses of a Spatially and Geochemically Confined Acid Mine Drainage Ecosystem Enabled by Community Genomics

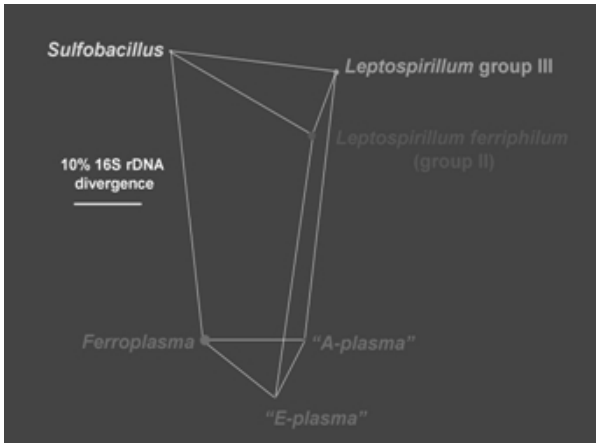
Gene W. Tyson, Philip Hugenholtz, and **Jillian F. Banfield** (jill@eps.berkeley.edu)

Department of Environmental Science, Policy and Management, Earth and Planetary Sciences, University of California, Berkeley

Subsurface acid mine drainage (AMD) ecosystems are ideal models for the genome-enabled study of microbial ecology and evolution because they are physically isolated from other ecosystems and are relatively geochemically and biologically simple.

We are using culture-independent genome sequencing of an AMD community from the Richmond mine, Iron Mountain, CA, to evaluate the extent and character of lateral gene transfer (LGT) within the community and to resolve microbial community function at the molecular level.

Microbial communities exist in several distinct habitats within the Richmond mine, including biofilms (subaqueous slime streamers and subaerial slimes) and cells attached directly to pyrite granules. All communities investigated to date by 16S rDNA clone libraries comprise only a handful of phylogenetically distinct organisms, typically dominated by the iron-oxidizing genera *Leptospirillum* and *Ferroplasma*. A *Leptospirillum*-dominated biofilm community was chosen for detailed analysis. 16S rDNA clone libraries and fluorescence *in situ* hybridization (FISH) using group-specific oligonucleotide probes indicated that the community is made up of only 6 prokaryotic populations (see Figure; the size of colored circles indicates 16S rDNA divergence within populations).



We analyzed initial community genome sequence data from a 3 Kb shotgun library of the biofilm to estimate the community genome size. This analysis used an implementation of the Lander-Waterman equation that took into consideration species abundances determined from the data and by FISH. Results indicate that each population is dominated by a single genome type. The conclusion was robust, even when assembly criteria were varied to improbable extremes and large uncertainties in population structure were included. However, more sequence data are needed to statistically validate the finding. Furthermore, the analysis is insensitive to genome types that occur in low abundance. The results suggest that reassembly of the dominant genomes of AMD community members will be tractable with a modest sequencing effort. The apparent population homogeneity may arise due to the specific characteristics of the AMD habitat or may be a widespread phenomenon in microbial ecosystems.

Similarity searches of the initial community genome sequence data revealed many genes consistent with the chemoautotrophic lifestyle of the community, including CO_2 fixation genes and nitrogen fixation genes. None of these key functional genes had close matches to genes from *Ferroplasma acidarmanus,* isolated from the Richmond Mine, the only relevant organism for which a complete genome is available. Thus, most of these genes likely come from *Leptospirillum*. Organism-resolved metabolic pathway information will be used to develop methods to monitor microbial activity in the environment.

LGT is thought to play a crucial role in the ecology and evolution of prokaryotes. The extreme conditions (pH < 1.0, molar concentrations of iron sulfate and mM concentrations of arsenic, copper and zinc, and elevated temperatures of up to 50° C) largely isolate the AMD community from most potential gene donors. Naked DNA, phage and prokaryotes native to neutral pH habitats do not persist at pH <1.0, precluding influx of genes by transformation, transduction and conjugation, respectively. However, prophage have been recognized in the *Ferroplasma* genome sequence and acidophilic phage have been detected in the biofilm community. Phage may be important vectors for gene exchange. We have initiated a collaboration to sequence the phage community to assess their diversity and enhance our ability to detect prophage in the prokaryote community genome data.

Comparative genome analyses indicate that *F. acidarmanus* and the ancestor of two acidophilic *Thermoplasma* species belonging to the Euryarchaeota have traded many genes with phylogenetically remote acidophilic *Sulfolobus* species (Crenarchaeota). The putatively transferred sets of *Sulfolobus* genes in *Ferroplasma* and the *Thermoplasma* ancestor are distinct, suggesting independent LGT events between organisms living in the same, and adjacent habitats. In both cases, however, the majority of transferred genes are involved in metabolism, particularly energy production/conversion and amino acid transport/metabolism. The lack of genes transferred from the (sequenced) genomes of other

prokaryotes is consistent with the hypothesis that extreme acidophiles have limited access to genes from organisms outside their ecotype. Interestingly, *Sulfolobus*, *Ferroplasma* and *Thermoplasma* are all bounded by a single tetraether-dominated membrane, which may facilitate conjugation. To date, no *Sulfolobus* species have been detected at Iron Mountain, suggesting two possibilities to explain the observed pattern of putatively transferred genes to *Ferroplasma* from *Sulfolobus*: 1) *Sulfolobus* is present at Iron Mountain but in regions currently inaccessible to sampling and/or 2) the transfers occurred prior to introduction of *Ferroplasma* into the current geological setting. Comparative analyses of the community genome data should improve the resolution of LGT in the community.

Ultimately, our goal is to develop an understanding of how acidophilic organisms evolved and function as communities to control acid mine drainage generation. The community genomics data are essential for this effort.