May 24, 2005

Final Progress Report
DE-FC-02-95ER61962
An Integrated Program in Microbial Genome Sequencing and Analysis
Claire M. Fraser, Ph.D. – Principal Investigator

The final progress report for this project contains information on nine microbial genome sequencing projects and two functional genomics projects that have been underway since the last report was submitted. As summarized below, work funded under this award has resulted in six publications, one submitted manuscript, and five manuscripts in preparation.

Genome Sequencing Projects

(1) *Carboxydothermus hydrogenoformans* Z-2901

Chemolithotrophic metabolism fuels primary production in many hydrothermal ecosystems, representing energy conservation strategies that may be independent of sunlight. In hot spring environments, chemolithotrophic metabolism may also complement phototrophic primary production in a variety of ways. Most autotrophic metabolic pathways in geothermal microbial communities depend on the presence of H2 gas in solution, to reduce O or S compounds, producing water or H2S. To better understand anaerobic carboxydotrophs that are capable of using CO as a carbon and energy source, producing gaseous hydrogen, we have completed the sequencing and analysis of the genome of the thermophilic bacterium *Carboxydothermus hydrogenoformans*. This species is a model system for the hydrogenogens, which are diverse organisms that grow anaerobically utilizing carbon monoxide as their sole carbon source and water as an electron acceptor, producing carbon dioxide and hydrogen as waste products. Analysis of the genome provides insight into how *C. hydrogenoformans* is able to grow so rapidly, despite the extremely low Gibbs free energy available to hydrogenogens. One likely contributing factor is the presence of five highly divergent carbon monoxide dehydrogenases. These five systems likely are responsible for this species' ability to grow on CO as its sole carbon and energy source. We have also identified what appears to be a minimal set of sporulation genes in this species. Using phylogenetic profiling, we have identified many uncharacterized gene families found in all known sporulating Firmicutes and not in any non-sporulating bacteria. A manuscript describing these results will be submitted in the next several weeks.

Frank T. Robb, Juan M.Gonzalez, Tatyana Sokolova, Stephen M. Techtmann, Nicolai Chernyh, Alexander Lebedinski, Luke J. Tallon, Kristie Jones, Martin Wu, Jonathan A. Eisen. In press. Primary Energy Metabolism in Geothermal Environments: The Role of Carbon Monoxide. In "1st Biannual Workshop on Geothermal Biology and Geochemistry in Yellowstone National Park," William Inskeep, Editor.

(2) *Acidothiobacillus ferrooxidans*

A. ferrooxidans is a key microorganism in biomining environments. It dominates such environments because of its ability to oxidize both iron and sulfur compounds, and shows unusually high tolerance to the low acidity typical of leaching operations. The genome sequence for this organism represents a single circular chromosome of 2,982,398 bp. We are comparing our annotation with that of a group in Chile that has been doing experimental work on the organism. A manuscript describing these results is in preparation.

(3) *Methylococcus capsulatus* Bath

Methanotrophs are ubiquitous bacteria that can use the greenhouse gas methane as a sole carbon and energy source for growth, thus playing major roles in global carbon cycles, and in particular, substantially reducing emissions of biologically generated methane to the atmosphere. Despite their importance, and in contrast to organisms that play roles in other major parts of the carbon cycle such as photosynthesis, no genome-level studies have been published on the biology of methanotrophs. We completed the first genome sequence from an obligate methanotroph, *Methylococcus capsulatus* (Bath), obtained by the shotgun sequencing approach. Analysis revealed a 3.3-Mb genome highly specialized for a methanotrophic lifestyle, including redundant pathways predicted to be involved in methanotrophy and duplicated genes for essential enzymes such as the methane monooxygenases. We used phylogenomic analysis, gene order information, and comparative analysis with the partially sequenced methylotroph *Methylobacterium extorquens* to detect genes of unknown function likely to be involved in methanotrophy and methylotrophy. Genome analysis suggests the ability of M. capsulatus to scavenge copper (including a previously unreported nonribosomal peptide synthetase) and to use copper in regulation of methanotrophy, but the exact regulatory mechanisms remain unclear. One of the most surprising outcomes of the project is evidence suggesting the existence of previously unsuspected metabolic flexibility in *M. capsulatus*, including an ability to grow on sugars, oxidize chemolithotrophic hydrogen and sulfur, and live under reduced oxygen tension, all of which have implications for methanotroph ecology. The availability of the complete genome of *M. capsulatus* (Bath) deepens our understanding of methanotroph biology and its relationship to global carbon cycles. We have gained evidence for greater metabolic flexibility than was previously known, and for genetic components that may have biotechnological potential.

Ward N, Larsen O, Sakwa J, Bruseth L, Khouri H, Durkin AS, Dimitrov G, Jiang L, Scanlan D, Kang KH, Lewis M, Nelson KE, Methe B, Wu M, Heidelberg JF, Paulsen IT, Fouts D, Ravel J, Tettelin H, Ren Q, Read T, DeBoy RT, Seshadri R, Salzberg SL, Jensen HB, Birkeland NK, Nelson WC, Dodson RJ, Grindhaug SH, Holt I, Eidhammer I,

Jonasen I, Vanaken S, Utterback T, Feldblyum TV, Fraser CM, Lillehaug JR, Eisen JA. (2004) Genomic insights into methanotrophy: the complete genome sequence of *Methylococcus capsulatus* (Bath). PLoS Biol. 2(10):e303.

(4) *Gemmata obscuriglobus* UQM 2246

G. obscuriglobus is a member of the planctomycete group of Bacteria. These organisms possess a unique combination of morphological and ultrastructural properties, including budding replication, the presence of crateriform structures of unknown function on the cell surface, a diverse range of extracellular appendages, and lack of the "universal" cell wall polymer peptidoglycan. In addition to their morphological and biochemical uniqueness, the planctomycetes are phylogenetically distinct from the rest of the Bacteria, occupying a separate branch of the 16S rRNA-based tree. In recent years, the planctomycetes have been found to be widely distributed and often numerically abundant in both aquatic and terrestrial environments. This group also includes the "missing lithotrophs" – organisms that can break down ammonia in wastewater anaerobically**;** this ecological niche has been postulated for many years, but the organisms performing this role have only recently been identified as planctomycetes. Other proposed environmental roles include degradation of chitin in marine systems, and the breakdown of toxic algal blooms. The random phase of this project has now been successfully completed, with 146,292 total sequences generated. Average sequencing success rate has been 82.9%, with average read lengths of 600 bases. The resulting 121,262 "good" sequences represent approximately 8-fold coverage of the genome. Remaining work on this project (not initially budgeted) includes closure and annotation; funds are being sought from a number of potential sources.

(5) *Geobacter sulfurreducens*

The complete 3.89 Mb genome sequence of *Geobacter sulfurreducens*, a delta-proteobacterium, reveals unsuspected capabilities, including evidence of aerobic metabolism, one-carbon and complex carbon metabolism, motility, and chemotactic behavior. The defining characteristic of this family is their ability to conserve energy to facilitate growth by oxidizing organic compounds to carbon dioxide and water using metals including uranium as electron acceptors. These abilities have resulted in considerable interest in their importance in the global cycles of these metals and carbon as well as their potential to act as agents in biorestoration efforts. These characteristics, coupled with the possession of many two-component sensors and many c-type cytochromes, reveal an ability to create alternative, redundant, electron transport networks and offer insights into the process of metal ion reduction in subsurface environments. As well as playing roles in the global cycling of metals and carbon, this organism clearly has the potential for use in bioremediation of radioactive metals and in the generation of electricity.

Methe BA, Nelson KE, Eisen JA, Paulsen IT, Nelson W, Heidelberg JF, Wu D, Wu M, Ward N, Beanan MJ, Dodson RJ, Madupu R, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Gwinn M, Kolonay JF, Sullivan SA, Haft DH, Selengut J, Davidsen TM, Zafar N, White O, Tran B, Romero C, Forberger HA, Weidman J, Khouri H, Feldblyum TV, Utterback TR, Van Aken SE, Lovley DR, Fraser CM. (2003) Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments. Science. 302:1967-9.

(6) *Colwellia psychroerythraea*

The completion of the 5,373,180 base pair (bp) genome sequence of the marine psychrophilic bacterium, *Colwellia psychrerythraea* 34H, a model for the study of life in permanently cold environments, reveals capabilities important to carbon and nutrient cycling, bioremediation, production of secondary metabolites and cold-adapted enzymes. From a genomic perspective, cold adaptation is suggested in several broad categories involving changes to the cell membrane fluidity, uptake and synthesis of compounds conferring cryotolerance, and strategies to overcome temperature-dependent barriers to carbon uptake. Modeling of three-dimensional protein homology from bacteria representing a range of optimal growth temperatures suggests changes to proteome composition that may enhance enzyme effectiveness at low temperatures. Comparative genome analyses suggest that the psychrophilic lifestyle is most likely conferred not by a unique set of genes but by a collection of synergistic changes in overall genome content and amino acid composition.

Barbara A. Methe, Karen E. Nelson, Jody W. Deming, Bahram Momen, Eugene Melamud, Xijun Zhang, John Moult, Ramana Madupu, William C. Nelson, Robert J. Dodson, Lauren M. Brinkac, Sean C. Daugherty, Anthony S. Durkin, Robert T. DeBoy, James F. Kolonay, Steven A. Sullivan, Liwei Zhou, Tanja M. Davidsen, Martin Wu, Adrienne L. Huston, Matthew Lewis, Bruce Weaver, Janice F. Weidman, Hoda Khouri, Terry R. Utterback, Tamara V. Feldblyum, and Claire M. Fraser. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. Proc. Natl. Acad. Sci. USA (submitted).

(7) *Dehalococcoides ethenogenes*

Dehalococcoides ethenogenes is the only bacterium known to reductively dechlorinate the groundwater pollutants, tetrachloroethene (PCE) and trichloroethene, to ethene. Its 1,469,720-base pair chromosome contains large dynamic duplicated regions and integrated elements. Genes encoding 17 putative reductive dehalogenases, nearly all of which were adjacent to genes for transcription regulators, and five hydrogenase complexes were identified. These findings, plus a limited repertoire of other metabolic modes, indicate that D. ethenogenes is highly evolved to utilize halogenated organic compounds and H2. Diversification of reductive dehalogenase functions appears to have been mediated by recent genetic exchange and amplification. Genome analysis provides

insights into the organism's complex nutrient requirements and suggests that an ancestor was a nitrogen-fixing autotroph.

Seshadri R, Adrian L, Fouts DE, Eisen JA, Phillippy AM, Methe BA, Ward NL, Nelson WC, Deboy RT, Khouri HM, Kolonay JF, Dodson RJ, Daugherty SC, Brinkac LM, Sullivan SA, Madupu R, Nelson KE, Kang KH, Impraim M, Tran K, Robinson JM, Forberger HA, Fraser CM, Zinder SH, Heidelberg JF. (2005) Genome sequence of the PCE-dechlorinating bacterium I. Science. 307:105-8.

(8) *Desulfovibrio vulgaris* Hildenborough

Sulfate-reducing bacteria (SRB) are a heterogeneous group of anaerobic prokaryotes found nearly ubiquitously in nature. Their importance derives not only from their environmental functions, but also their economic roles; for example, anaerobic corrosion of steel, sulfide contamination of petroleum and paper products, and their potential for toxic metal remediation. The SRB are known for their metabolic interactions with metals and are notorious for their involvement in biocorrosion. These bacteria indirectly cause metal ion precipitation because the sulfide generated from sulfate reduction forms insoluble precipitates with a number of metals. Additionally, the discovery that SRB can also deliver electrons to a number of toxic metals, thereby altering their solubility, has revealed a potentially useful mechanism for the bioremediation of metal ion contaminants from anaerobic sediments. Among the metal ions known to be reduced by the SRB are uranium, technetium, chromium, and arsenic (Lovley, 1993), with most becoming less soluble in the reduced form. Before environmental bioremediation can be productively applied, much needs to be learned about the delivery of electrons to the metals and the regulation and specificity of the reductases.

Desulfovibrio vulgaris Hildenborough is a model organism for studying the energy metabolism of sulfate-reducing bacteria (SRB) and for understanding the economic impacts of SRB, including biocorrosion of metal infrastructure and bioremediation of toxic metal ions. The 3,570,858 base pair (bp) genome sequence reveals a network of novel c-type cytochromes, connecting multiple periplasmic hydrogenases and formate dehydrogenases, as a key feature of its energy metabolism. The relative arrangement of genes encoding enzymes for energy transduction, together with inferred cellular location of the enzymes, provides a basis for proposing an expansion to the 'hydrogen-cycling' model for increasing energy efficiency in this bacterium. Plasmid-encoded functions include modification of cell surface components, nitrogen fixation and a type-III protein secretion system. This genome sequence represents a substantial step toward the elucidation of pathways for reduction (and bioremediation) of pollutants such as uranium and chromium and offers a new starting point for defining this organism's complex anaerobic respiration.

Heidelberg JF, Seshadri R, Haveman SA, Hemme CL, Paulsen IT, Kolonay JF, Eisen JA, Ward N, Methe B, Brinkac LM, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS, Madupu R, Nelson WC, Sullivan SA, Fouts D, Haft DH, Selengut J, Peterson JD,

Davidsen TM, Zafar N, Zhou L, Radune D, Dimitrov G, Hance M, Tran K, Khouri H, Gill J, Utterback TR, Feldblyum TV, Wall JD, Voordouw G, Fraser CM. (2004) The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. Nat Biotechnol. 22:554-9. Epub 2004 Apr 11.

(9) *Thermotoga neapolitana*

The genome of the thermophilic bacterium *Thermotoga maritima* (Huber, Langworthy et al. 1986) was sequenced in 1999 (Nelson, Clayton et al. 1999). The closely related *Thermotoga neapolitana* was isolated near Naples (Jannasch, Huber et al. 1988) across the bay from where *T. maritima* was isolated in Vulcano, Italy in 1986 (Huber, Langworthy et al. 1986). This bacterium also grows on a range of carbohydrates including ribose, glucose, sucrose maltose and lactose. Many initial studies on this bacterium have focused on carbohydrate utilization patterns as well as more recently hydrogen production potential. In addition to the insights to be gained on carbohydrate metabolism in these organisms, it is anticipated that the genome of *T. neapolitana* may provide additional evidence for gene transfer events by members of this lineage (Nesbo, L'Haridon et al. 2001; Nesbo, Nelson et al. 2002). In an attempt to further characterize this phenomenon in members of this lineage, we recently completed a whole genome sequencing project for *T. neapolitana* strain NS-ET which is also the type strain of this species. The genome of *T. neopolitana* NSE was sequenced by the random shotgun method, with cloning, sequencing and assembly as described previously.

The genome of *T. neapolitana* is 1 884 562 bp long with an average GC content of 46.9%. The genome encodes for a total of 1906 ORFs with an average read length of 939 bp. 70.9 % of the predicted ORFs could be assigned to a role category, with the remainder being conserved hypothetical proteins (19.4%), or proteins of unknown function (8.1%). 43 ORFs (2.3% of total) were unique to *T. neapolitana*. In comparison, the genome of *T. maritima* is 1 860 725 bp in length, with an estimated 1876 ORFs and an average GC content of 46.3%. The genomes of these two species are essentially linear (Figure 1A-C), but there are four major rearrangements in the genome of *T. neapolitana* when compared to that of *T. maritima*. All four major inversions are flanked by repetitive sequences that have been associated with CRISPRs (clustered regularly interspaced palindromic repeats) proteins in the genomes of both *T. neapolitana* and *T. maritima*. CRISPRS proteins and the associated repeats have been identified in intergenic spacer regions of a number of microbial species for who complete genome sequences are available (Nelson, Clayton et al. 1999; Jansen, Embden et al. 2002; Schouls, Reulen et al. 2003). Each CRISPR locus is characterized by a repeated DNA sequence that is spaced by unique intervening sequence. The role of these elements in microbial genomes is unknown, but our comparative genome analysis of the two *Thermotogas* presented here suggests that they may target tRNAs, are associated with the major DNA rearrangements between these two species, and may in fact be previously uncharacterized mobile elements. A possible role in DNA mobility (and lateral gene transfer) was proposed in the original genome paper of *T. maritima* as the DNA repeat that is characteristic of the CRISPR element was found in close proximity to predicted archaeal-islands in a few

instances (Nelson, Clayton et al. 1999; Jansen, Embden et al. 2002; Schouls, Reulen et al. 2003)

As a result of the *T. neapolitana* genome study, there are currently three manuscripts in preparation:
Comparative and functional analysis of *Thermotoga neopolitana*. Karen E. Nelson, Emmanuel Mongodin, Sean Daugherty, Ioana Hance, Robert T. DeBoy, and Claire M. Fraser.

CRISPRs and tRNA genes as features for shuffling genomes in the Thermotogales. Robert T. DeBoy, Emmanuel Mongodin, Joanne Emerson, and Karen E. Nelson.

Definition and Characterization of Distinct Subtypes of Clustered Regularly Interspaced Palindromic Repeats (CRISPR) Associated Proteins Among Prokaryotic Genomes. Daniel H. Haft, Jeremy Selengut, Emmanuel F. Mongodin, Karen E. Nelson (will be submitted week of May 23rd, 2005).

Functional Genomics Projects

(1) Expression analysis of simple carbohydrate metabolism in *Thermotoga maritima*

The thermophilic bacterium, *Thermotoga maritima*, is a heterotrophic organism capable of metabolizing both simple carbohydrates and complex carbohydrates such as cellulose and xylan. As a first step towards identifying genes and metabolic pathways involved in carbohydrate metabolism by *T. maritima*, we have completed an effort focused on using *T. maritima* DNA microarrays to identify genes required for metabolism of the simple carbohydrates; glucose, lactose and maltose in a chemostat. A much larger number of genes changed expression in cells grown on lactose than on maltose, each relative to genes expressed in cells grown on glucose. Genes encoding putative oligopeptide transporters were often coregulated with adjacent glycosidase-encoding genes. Genes encoding enzymes catalyzing NADH oxidation were up-regulated on both lactose and maltose. Genes involved in iron and sulfur metabolism were differentially expressed in response to lactose. These data help define the sets of coregulated genes and suggest possible functions for their encoded products.

Nguyen TN, Ejaz AD, Brancieri MA, Mikula AM, Nelson KE, Gill SR, Noll KM. (2004). Whole-genome expression profiling of *Thermotoga maritima* in response to growth on sugars in a chemostat. *J Bacteriol*. 186:4824-8.

(2) Transcriptional profiling of *Methanococcus jannaschii*

Temperature shock of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* from its optimal growth temperature of 85 degrees C to 65 degrees C and 95 degrees C resulted in different transcriptional responses characteristic of both the direction of shock (heat or cold shock) and whether the shock was lethal. Specific outcomes of lethal heat

shock to 95 degrees C included upregulation of genes encoding chaperones, and downregulation of genes encoding subunits of the H(+) transporting ATP synthase. A gene encoding an alpha subunit of a putative prefoldin was also upregulated, which may comprise a novel element in the protein processing pathway in *M. jannaschii*. Very different responses were observed upon cold shock to 65 degrees C. These included upregulation of a gene encoding an RNA helicase and other genes involved in transcription and translation, and upregulation of genes coding for proteases and transport proteins. Also upregulated was a gene that codes for an 18 kDa FKBP-type PPIase, which may facilitate protein folding at low temperatures. Transcriptional profiling also revealed several hypothetical proteins that respond to temperature stress conditions.

Boonyaratanakornkit BB, Simpson AJ, Whitehead TA, Fraser CM, El-Sayed NM, Clark DS. (2005). Transcriptional profiling of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* in response to lethal heat and non-lethal cold shock. Environ Microbiol. 7:789-797.