Office of Science
Financial Assistance
Funding Opportunity Announcement
DE-PS02-07ER07-14

New Genomic Strategies and Technologies for Studying Complex Microbial Communities and Validating Genomic Annotations

The Office of Biological and Environmental Research (BER) of the Office of Science (SC), U.S. Department of Energy (DOE), hereby announces its interest in receiving applications for research that supports the Genomics: GTL research program (http://genomicsgtl.energy.gov/). In this Notice, two areas of research applications are sought. The first is concerned with characterization of microbial communities having high priorities for DOE missions in bioenergy production, carbon cycling, and environmental remediation. The science of microbial ecology will be advanced by understanding the activities, composition (at multiple scales of resolution), distribution, diversity, relative abundance, and interactions of the microorganisms in relevant microbial communities. The second area addresses the mismatch between genomic and metagenomic DNA sequencing capabilities and the experimental testing of function annotations derived from the newly displayed sequence. Novel approaches are invited to better characterize microbial communities based on genomic information and/or to accelerate experimental validation of genome, gene set, and metagenomic annotations.

PREAPPLICATIONS

Potential applicants are required to submit a brief preapplication, referencing **Program Notice DE-PS02-07ER07-14 for receipt by DOE by 4:30 p.m., Eastern Time, January 18, 2007.** Preapplications will be reviewed for conformance with the guidelines presented in this Notice and suitability in the technical areas specified in this Notice. A response to the preapplications encouraging or discouraging formal applications will be communicated to the applicants by **January 26, 2007.** Applicants who have not received a response regarding the status of their preapplication by this date are responsible for contacting the program to confirm this status.

Only those preapplicants that receive notification from DOE encouraging a formal application may submit full applications. **No other formal applications will be considered.**

Potential applicants **must** submit a brief preapplication that consists of no more than three pages of narrative stating the research objectives, describing the technical approach(s), and identifying the proposed team members and their expertise. The intent in requesting a preapplication is to save the time and effort of applicants in preparing and submitting a formal project application that may be inappropriate for the program. Preapplications will be reviewed relative to the scope and research needs as outlined in the summary paragraph and in the SUPPLEMENTARY INFORMATION. The preapplication should identify, on the cover sheet, the title of the project,

the institution or organization, principal investigator name, telephone number, fax number, and e-mail address. No budget information or biographical data need be included, nor is an institutional endorsement necessary. Preapplications referencing the first aim in Program Notice DE-PS02-07ER07-14 should be sent as a text file without attachments or a single PDF file attachment via e-mail to: **genomicsGTL@science.doe.gov** with **"Preapplication DE-PS02-07ER07-14 Microbial Communities Lastname Institution"** as the subject. Preapplications referencing the second aim in Program Notice DE-PS02-07ER07-14 should be sent as a text file without attachments or a single PDF file attachment via e-mail to: **genomicsGTL@science.doe.gov** with **"Preapplication DE-PS02-07ER07-14 Annotation Lastname Institution"** as the subject. No FAX or mail submission of preapplications will be accepted.

APPLICATION DUE DATE: March 20, 2007, 8:00 pm, Eastern Time

Applications must be submitted using Grants.gov, the Funding Opportunity Announcement can be found using the CFDA Number, 81.049 or the Funding Opportunity Announcement number, DE-PS02-07ER07-14. Applicants must follow the instructions and use the forms provided on Grants.gov.

FOR FURTHER INFORMATION CONTACT:

For the first aim of this notice, contact Dr. Daniel Drell,
telephone: (301) 903- 4742, E-mail: Daniel.drell@science.doe.gov.

For the second aim of this notice, contact Dr. Arthur Katz,
telephone: (301) 903-4932, E-mail: Arthur.katz@science.doe.gov

SUPPLEMENTARY INFORMATION:

The Genomics:GTL Program is a systems biology research program addressing key DOE missions and national priorities including: developing abundant sources of clean energy, controlling greenhouse gases like carbon dioxide (a key factor in global climate change), and helping to clean up past contamination of the environment. Microorganisms are the largest reservoir of genetic and biochemical diversity on Earth and they and their resulting communities mediate many natural processes of direct mission relevance to the DOE Genomics:GTL Program, including: 1) biodegradation and bioconversion of complex polymers into simpler component substances, many of which offer great promise as sources of potential energy sources and fuels (e.g. ethanol, butanol, hydrogen, methane, etc.); 2) management of the fate and transport of contaminants comprising or containing heavy metals, radionuclides, or chlorinated solvents (e.g. uranium, chromium, plutonium, mercury, carbon tetrachloride, trichloroethylene, perchlorethylene) in subsurface areas at DOE sites; and 3) management of the flow and fate of carbon (including carbon dioxide) through terrestrial and marine environments with consequent impacts on atmospheric CO2 concentrations.

Studies of single microbes have revealed some of the relevant biochemistries but do not capture the properties of natural microbial communities that mediate these processes. Recent

examinations of microbial communities have revealed that hard-to-culture microbes make up more than 99% of many natural microbial communities and thus much of our present knowledge is based on a very small subset of their component members. DNA isolated directly from environmental samples is a tremendous resource for examining, in a more comprehensive and representative way, the structure and functioning of microbial communities but analyses of "metagenomic" DNA are not well advanced. The science of microbial ecology pertinent to DOE missions in bioenergy, fate and transport of contaminants, and carbon cycling will be advanced by understanding the activities, compositions, distribution, diversity, relative abundance, and interactions of the full range of microorganisms in relevant communities. A challenge to achieving this objective, however, is the difficulty in characterizing the complexity of microbial communities in nature and the ways they may change over time. Recent "metagenomic" sequencing surveys have indicated the presence of extraordinary numbers (in the tens of *millions*) of previously unknown genes. Thus, new strategies and technologies are needed to help define and assess the repertoire of metabolic capabilities as embodied in the collective genomic sequence of a microbial community.

In addition, there is a mismatch between genome DNA sequencing capabilities and the validation of proposed functional assignments derived from sequence interpretation. Gene structure recognition within genome sequence uses any of several software packages. A subsequent functional annotation is then derived from comparisons of the resulting gene models with those from other organisms previously sequenced and annotated. For a microbial species not previously sequenced, about a third of the putative genes lack homologues from which to predict function. For those sequences having homologues, the initial functional assignments may be questionable due to errors from a variety of sources. Some initial annotations in the public databases are simply not correct. The deduced gene model may not be correct due to erroneous DNA sequence calls or novelties in gene structure not yet recognizable by the software. Homology based sequence assignments become less reliable as the phylogenetic distance between compared species grows. Examples are known of proteins with very similar structures but different functions. Similarly, examples exist of proteins with dramatically different structures but very similar functions. Some proteins have multiple functions. While human curation can correct some annotation errors, the flood of new sequences is swamping curatorial efforts. Ultimately, there is no substitute for high-throughput experimental methods to test gene annotations which are crucial to practical applications. The aim of this part of this call is to improve annotation testing for genes having greatest impact on DOE projects.

This leads directly to the two main thrusts of this Notice: 1) to develop technologies for characterization and functional analyses of microbial communities involved in DOE mission relevant activities and 2) to further stimulate innovative experimental technologies and approaches to test predicted gene/protein function(s) including those entirely lacking homologues.

1) New technologies and strategies for characterization and functional analyses of microbial communities involved in DOE mission relevant activities are encouraged that will:

- Use genome sequences from microbial communities to identify, in high-throughput, the mission relevant genes, metabolic pathways, regulatory networks and proteins needed for (or that impact) survival, growth and adaptation to the environment and especially the metabolic and physiologic reactions carrying out degradations of cellulose (and cellulose-linked compounds) or other bioenergy relevant processes, carbon cycling, or management of the fate and transport of high-priority contaminants. In particular, novel and innovative methods and approaches are needed to identify community composition and genetic diversity in microbial communities of interest to the DOE. Methods for conceptual reconstruction, from sequence data, of essential metabolic and DOE mission relevant pathways are of interest. It is critical to understand the relationships between genetic diversity, community diversity, and microbial community functioning to provide insights into community "functional regulation" as a result of microbial community composition. Better approaches to "binning" sequences of novel genes into putative (operationally defined) "species" and pathways are needed. In responding to this part of this notice, applicants should take particular care to define the extent of the "community" they intend to study.

  Key technologies needed to achieve these goals include, but are not limited to:

- New highly parallel comparative approaches that allow unique microbial community DNA fragments to be identified and the community to be characterized in automated high-throughput ways.
- Novel technologies and approaches for defining the patterns of expression and functions of genes from, and co-incident with, microbial communities under different environmental conditions.

2. Develop and apply high-throughput innovative technologies and approaches to improve gene/protein functional assignments:

Genome sequence is available for a rapidly growing number of individual microbes, plants, and microbial communities. This sequence information is made available to the public, commonly with initial gene identifications and annotations determined by computational analyses. A key challenge is to efficiently test and improve the annotations of these putative genes and their potential function(s).

Currently, annotation of genes largely relies on computational algorithms to recognize genes within DNA sequence, and then carry out a homology-based search of the sequences of genes already characterized to infer possible function. However, relying on computational algorithms alone for annotation has substantial limitations. It has been increasingly clear that there is no substitute for experimental methods to test and validate gene annotation.

Consequently, this element addresses this imperative by supporting the development of improved technologies for high throughput experimental validation. Genomics:GTL seeks improvements to high throughput testing of gene annotation for microbes and microbial gene sets of interest to DOE missions in bioenergy, carbon cycling, and bioremediation. The Genomics:GTL Program seeks creative and innovative thinking and strategies that can propose new approaches as well as

the coupling, where appropriate, of a combination of new and/or existing technologies into flexible and efficient systems for experimental annotation.

High throughput innovative approaches to functional analyses can use a variety of new and existing formats and technologies, individually or in combination, that can include but are not limited to, physiological assays, structural and physical measurements using purified proteins, and in vivo phenotypic analyses of microbes. Microarrays, mass, force and electromagnetic spectroscopies, single cell analyses, phage display approaches, and/or other structural characterizations are among the possible techniques that may be used.

While the ultimate goal of this element is to develop experimental techniques to rapidly validate annotations, an initial test of these techniques would be to identify and validate genes and gene sets from genomic and metagenomic data that are part of networks and pathways that carry out functions of particular interest to DOE. Possible targets could include activities mediated by cytochromes, rhodopsins, dehydrogenases, cellulases, dehalogenases, etc.

Improvements to experimental technologies and strategies are thus sought to:

> Develop and/or innovatively apply techniques that can rapidly, with a high degree of confidence, improve gene/protein functional assignments. **These approaches should strive to be compatible with other complementary tools for gene/protein functional assignments.**

Potential approaches to the goal of high throughput experimental annotation include but are not limited to:

- The development of systems that combine protein and protein domain production with high throughput functional testing,
- The development of systems that combine genetic and reverse genetic approaches with high throughput functional testing.

***Microbes of Interest to DOE*.** The focus of Genomics:GTL is on nonpathogenic microbes (including fungi) and some plants that are directly relevant to DOE mission needs in energy (biomass conversion to energy sources, carbon sequestration, and the global carbon cycle both terrestrial and ocean) or the environment (fate and transport of metals, radionuclides, and certain organics at DOE sites). When possible, research within this solicitation should take advantage of and focus on microbes whose complete DNA sequence is already known or microbial communities of interest to, directly relevant to, or that would contribute substantially to an ability to address DOE mission needs. **Applicants should identify proposed high throughput DNA sequencing needs, if any, in their application; subject to merit review, accepted applications can expect sequencing needs to be satisfied at the DOE-Joint Genome Institute (DOE-JGI) at no cost following a technical review at the DOE-JGI.** Applicants should also provide a clear, scientifically justified description for their choice of microbe(s) in the context of DOE mission needs as outlined above.

***Data and Other Results*.** Microbial DNA sequence data will be publicly released according to the "Data Release Requirements: Microbial Genome Sequencing Projects" (

http://www.jgi.doe.gov/sequencing/collaborators/datarelease.html). Data and results that are generated through these investigations that are appropriate to share with the broader community should be provided in timely, open, and machine-readable format where possible. Funded investigators are expected to contribute to and participate with the GTL working group on data management, and to adhere to the group's consensus on data sharing.

The Genomics:GTL program supports a combination of large, well integrated, multidisciplinary research teams and smaller, focused research projects. This solicitation will support smaller, focused research projects to develop new technologies, research strategies, or research resources needed by the Genomics:GTL program.

Information on the research projects currently funded by the Genomics: GTL program and a description of project goals and overall program organization can be found at: http://genomicsgtl.energy.gov/).

Other useful web sites include:

DOE Joint Genome Institute (JGI) Microbial Web Page - http://genome.jgi-psf.org/mic_cur1.html;

DOE Joint Genome Institute Sequencing projects: http://www.jgi.doe.gov/sequencing/seqplans.html

Microbe Project Interagency Working Group Home Page: http://www.microbeproject.gov/

Genomes On Line Database: http://www.genomesonline.org/

Posted on the Office of Science Grants and Contracts Web Site
November 24, 2006.