GTL P.I. Workshop, Washington, D.C.
Breakout Group: Environmental Genomics
Monday, March 1, 2004

Breakout Leaders: Craig Venter (Institute for Biological Energy Alternatives) and Eddy Rubin (Joint Genome Institute)

Discussion Topics

Informatics: The group discussed extensively the need for databases to capture all information associated with metagenomic projects. Although GenBank can accommodate some of the data, it is not set up to receive the full array that needs to be linked to such projects, including the following:

- Methods of sample collection. DNA will not be cultured organisms but rather isolated from filters or from soil samples. Methods of soil collection and DNA preparation will impact the data.

- Environmental index information describing the sampling site: longitude, latitude, time of day, depth, temperature, and geochemical features. With the goal of someday being able to take a systems biology approach to various environments, scientists consider essential the ability to link these environmental sequences with a wealth of other identifiers.

- Capture of polymorphism data in environmental sequence. Important information is contained in all sequencing reads, so data should be kept rather than assembled into a consensus sequence.

Action Item: Organize a meeting of people involved in environmental genomics, including generators of environmental data, users, database manufacturers, and representatives from private industry and public and private funding agencies. Attendees pointed out the limited number of data-generating facilities, so establishing standards for study participants would be quite feasible. The potential constricting impact of developing such standards at this early stage also was raised.

Technologies for Environmental Genomics: The need to facilitate effective environmental samplings was discussed, along with ways to normalize sequencing to avoid wasting efforts on predominant organisms and missing less frequent organisms. Needed technologies include single-cell amplification, physical means such as cell sorting for separating microbes, high-throughput culturing, expression arrays to monitor samples, and efficient algorithms for assembling and annotating metagenomic data.

Action Item: Support technologies to facilitate various approaches to meaningful and cost-effective metagenomic studies.