

Systems Biology Research Symposium

Oral Presentation Session

Grand Ballroom
Tuesday, June 5th
7:00-8:30pm

A Data Analysis and Coordination Center for the Human Microbiome Project

Jennifer Wortman¹, Michelle Giglio¹, Amy Chen², Victor Felix¹, Konstantinos Mavrommatis³, Heather Creasy¹, Todd DeSantis², Clark Santee², Yvette Piceno², Shariff Osman², Joshua Orvis¹, Jonathan Crabtree¹, Micah Hamady⁴, Justin Kuczynski⁴, Rob Knight⁴, Gary Andersen², Nikos Kyrpides³, Victor Markowitz², and Owen White¹

¹Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; ²Lawrence Berkeley National Laboratory, Berkeley, CA; ³Joint Genome Institute, Walnut Creek, CA; ⁴University of Colorado at Boulder, Boulder, CO
Presenter's email address: mgiglio@som.umaryland.edu

The Human Microbiome Project (HMP) is an NIH Roadmap initiative that aims to collect and analyze unprecedented amounts of sequence information from microbial communities found in and on the human body. There is abundant and growing evidence that changes in microbial community composition are highly correlated with human health and disease. Efforts are underway to determine if such changes are the result of particular human diseases or perhaps a contributing cause. To gain insight into this question, the HMP has undertaken two main areas of effort: sequence 1000 reference genomes that live in or on the human body and sequence metagenomic samples from five different body sites collected in parallel from healthy subjects and those with disease. Initially, four large sequencing centers have begun the work of sequencing the 1000 reference strains. Subsequently, centers will be funded to carry out metagenomic sequencing from various sites with subjects suffering from various conditions. This project will generate unprecedented amounts of sequence data, annotation information, and metadata about subjects and strains. The analysis of this data requires the ability to collect, integrate, and standardize information of different types and from different sources. Responsibility for these activities falls on the HMP Data Analysis and Coordination Center (DACC). Successful data integration and standardization will rely on the use of controlled vocabularies, the application of quality control measures, and the development of standard operating procedures. The DACC will provide multiple analysis services to the research community including data query, comparative genomics, 16S rRNA analysis, and phylogenetic analysis. The DACC will also engage in extensive training and outreach. All information and analyses produced from the HMP will be available on a comprehensive web resource. The web presentation toolsets for the DACC will be based on those of the Integrated Microbial Genomes resource for both single genomes and metagenomes (IMG and IMG/m). The HMP DACC can be found at <http://hmpdacc.org>.

Key words: metagenomics, human microbiome, microbe-host interactions