

IMG/M: An Experimental Metagenome Data Management and Analysis System

Natalia Ivanova³, Victor M. Markowitz¹, Frank Korzeniewski¹, Krishna Palaniappan¹, Ernest Szeto¹, Inna Dubchak², Athanasios Lykidis³, Iain Anderson³, Konstantinos Mavrommatis³, Victor Kunin⁴, Hector Garcia Martin⁴, Philip Hugenholtz⁴, and Nikos C. Kyrpides³

¹Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, USA, ² Genomics Division, Lawrence Berkeley National Laboratory, USA, ³Genome Biology Program, Joint Genome Institute, USA, ⁴Microbial Ecology Program, Joint Genome Institute, USA.

Environmental microbial community (microbiome) genome (*metagenome*) analysis is expected to lead to advances in environmental cleanup, agriculture, alternative energy production, and treatment of human diseases. The application of shotgun sequencing to microbiome samples has enabled the study of metagenomes involving previously uncultured and unculturable organisms.

We have developed an experimental metagenome data management and analysis system, IMG/M, based on the Integrated Microbial Genomes (IMG) system. In addition to IMG's isolate genomes, IMG/M includes metagenome sequence data generated from several environmental microbiome samples which comprise a representative set in terms of species diversity, abundance of dominant organism(s) and sequencing depth. The IMG/M data analysis tools are the result of extending the IMG data analysis tools to handle metagenome data. Additional tools allow examining profiles of functional annotations across microbial communities and isolate organisms of interest, and analyzing strain-level heterogeneity within a species population in metagenome data.

IMG/M shows that although data processing of metagenome sequences needs to overcome numerous challenges due to the complex nature and inherent incompleteness of the data, successful analysis can be conducted in the context of a comprehensive data management and analysis system that provides support for data review and revision. IMG/M has been used for completing the analysis of biological phosphorus removing (EBPR) sludge communities and for studying the metagenomes of several key microbial communities recently sequenced by JGI, including the hydrogen-producing consortium colonizing the termite hindgut. IMG/M will be demonstrated from both data management and analytical perspectives. IMG/M is available at: <http://img.jgi.doe.gov/m>.

This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research; the University of California; and Lawrence Berkeley, Lawrence Livermore, and Los Alamos National Laboratories.

IMG/M software and data are free and available without restrictions on use or access.