

Data acquisition and coordination key to human microbiome project

June 9 2010



Phylogenetic analysis of 16S Ribosomal DNA sequences with Human Microbiome Project microbes highlighted in blue shows the distribution of these human symbionts around the microbial tree of life. Phylum are separated by color as follows: yellow, Actinobacteria; dark green, Bacteroidetes; light green, Cyanobacteria; red, Firmicutes; cyan, Fusobacteria; dark red, Planctomycetes; gray, Proteobacteria; magenta, Spirochaetes; light pink, TM7; tan, Tenericutes. Credit: Image courtesy of Human Microbiome Project

At birth, your body was 100-percent human in terms of cells. At death, about 10-percent of the cells in your body will be human and the remaining 90-percent will be microorganisms. That makes you a "supraorganism," and it is the interactions between your human and

microbial cells that go a long way towards determining your health and physical well-being, especially your resistance to infectious diseases.

To learn more about the community of symbiotic microbes that outnumber our own somatic and [germ cells](#) by a 10:1 ratio, the National Institutes of Health (NIH) in 2008 launched the Human Microbiome Project (HMP) - a microbiome is the full complement of microorganisms populating a supraorganism. The goal of the HMP is to sequence the genomes of 1,000 or more of these microbial species and assemble the information in a "project catalog" as a reference for future investigations. The project catalog is housed at the HMP [Data Acquisition](#) and Coordination Center (DACC), which was created and is maintained by researchers with the U.S. Department of Energy's Lawrence Berkeley National Laboratory (Berkeley Lab).

"The HMP project catalog is a unique worldwide resource," says molecular biologist Nikos Kyrpides of Berkeley Lab's Genomics Division, who heads the [Genome Biology](#) and Metagenomics Programs for the Joint Genome Institute (JGI) and is the co-principal investigator of the DACC. "It has a central role in the HMP, not only in maintaining the list and status of over 1,400 individual human microbiome projects, but also as a data managements system for the metadata associated with these projects, such as information on the microbial isolation sites and the sites in the human body where these microbes can be found, and information on the phenotypic properties of these microbes."

At JGI, Kyrpides oversees projects such as GenePRIMP, a highly rated quality control program for genome sequencing, and GOLD, the Genomes On-Line Database. GenePRIMP stands for "Gene PRediction IMprovement Pipeline, and it consists of a series of computational units that can be used to significantly improve the overall quality of the predicted genes in any sequenced genome. The results identify gene-calling errors such as potentially incorrect gene start and end positions,

large overlaps between genes, and fragmented or missed genes. GOLD provides comprehensive information on genome sequencing projects, including metagenomes and metadata from around the world. The HMP project catalog is powered by the GOLD database and provides a specialized user interface by which the data stored in GOLD can be read.

The other co-principal investigator of the DACC is Victor Markowitz who heads Berkeley Lab's Biological Data Management and Technology Center in the Computational Research Division, and also serves as the Chief Informatics Officer and Associate Director at JGI. Markowitz oversees the development and maintenance of the Integrated Microbial Genomics with Microbiome samples (IMG/M) system, which provides comparative analysis tools for the study of metagenomes - the collective genetic material of a given microbiome. First released in 2006, IMG/M contains millions of annotated microbial gene sequences, recovered from wild varieties of microbial communities. IMG/M is now being applied to the HMP.

"Resources such as GenePRIMP, GOLD and IMG/M are among the best in the world when it comes to providing comparative analysis tools for microbial genomes and metagenomes," Markowitz says. "As the HMP moves forward, these resources will provide support for the annotation and analysis of HMP datasets, in particular via the metagenome annotation pipeline at JGI and a HMP specific version of the IMG/M system."

The first 178 reference microbial genomes have now been analyzed and catalogued by the HMP. The results were published in the journal *Science* in a paper titled, "A Catalog of Reference Genomes from the Human Microbiome."

In this paper, HMP researchers report comparing data from the

sequenced reference genomes to human metagenomic data in the public domain to identify proteins, determine gene functionality and link metagenomic data to individual [microbial species](#). From an analysis of 547,968 predicted proteins, the HMP researchers report 29,987 unique proteins, which suggests a far greater diversity in the human microbiome than previously suspected.

"The Science paper is a milestone in the human microbiome research with the release to the public of 178 finished or high quality draft genomes from organisms isolated from various sites in the human body," says Kyrpides. "It signals the beginning of a much larger effort that aims to provide a more comprehensive genetic catalog of the microbes living in the human body. The impact of understanding what is the normal microbial flora, what is its core genetic content, and how perturbations of the normal microbial flora of the [human body](#) can shift from protecting our bodies into causing diseases will eventually be enormous."

Kyrpides, Markowitz and their colleagues at the DACC are playing a critical role in fulfilling an NIH call for development of common sequencing and annotation standards that have not existed before. Lack of common language and a clearing house for genome data have been among the most daunting problems in genomics research.

Says Markowitz, "The greatest challenge ahead will be handling hundred of metagenomic datasets generated as part of the HMP, which will represent several orders of magnitude more data than the datasets presented in the current paper. We need to develop novel analysis and visualization methods to handle this massive increase in data."

Adds Kyrpides, "New sequencing technologies and our ability to generate orders of magnitude more data compared to only a year or two ago are changing the field entirely, and are mandating a social shift among the scientists involved to a more collaborative rather than

competitive spirit. None of us can provide solutions alone any more, and joint efforts such as the HMP are the only way we'll succeed."

Provided by Lawrence Berkeley National Laboratory

Citation: Data acquisition and coordination key to human microbiome project (2010, June 9)
retrieved 2 May 2024 from

<https://phys.org/news/2010-06-acquisition-key-human-microbiome.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--