

Human Microbiome Project Data Available to Research Community

Baltimore, Maryland

May 23, 2011

The Human Microbiome Project (HMP), funded through the National Institutes of Health Common Fund's Roadmap for Medical Research, is now releasing reads and assembled sequences from whole metagenome shotgun sequencing of 690 microbiomes and ~ 72 million reads from targeted 16S sequencing of 5034 microbiomes from healthy human subjects for use by the scientific community via the [HMP Data Analysis and Coordination Center \(DACC\)](#), which is located at the Institute for Genome Sciences (IGS) at the University of Maryland School of Medicine.

Since its launch in 2008, the Human Microbiome Project has generated a huge volume of sequence, annotation and metadata. The HMP DACC is charged with the task of organizing this information and facilitating its use in analysis by researchers outside the HMP.

Now, on behalf of the HMP, the DACC is releasing both processed reads and assemblies from whole metagenome shotgun sequencing for an initial group of 690 metagenomes from a subset of >17,000 samples collected from 300 healthy human volunteers. Fifteen body habitats are represented in this collection. Assembly was carried out by the HMP Assembly Working Group. The DACC is in the process of generating a gene index of all proteins predicted from these assemblies which will be made available to the public upon completion. In addition ~72 million reads corresponding to deconvoluted, trimmed 16S sequences from 5034 samples are also available. 16S variable region V3-5 was sequenced for all 5034 samples, with variable regions V1-3 and V6-9 also sequenced for subsets of the samples. 18 body sites are represented in this 16S collection.

Sampling of the subjects was done at the Baylor College of Medicine and Washington University and sequencing was carried out by Washington University, Baylor College of Medicine, the J. Craig Venter Institute, and the Broad Institute.

Owen White, PhD, Director of Bioinformatics at the Institute for Genome Sciences, Professor of Epidemiology and Preventive Medicine at the University of Maryland School of Medicine, and Principal Investigator of the DACC explained, "These data provide the research community with a key resource for their human metagenomic studies, facilitating research into complex human diseases."

The whole metagenome read and assembly data is available at www.hmpdacc.org/HMASM/.

The 16S reads are available at www.hmpdacc.org/HM16STR/.

More information about the Human Microbiome Project is available at www.nihroadmap.nih.gov/hmp/.

About IGS

The Institute for Genome Sciences (IGS) at the University of Maryland School of Medicine is an international research center dedicated to advancing the use of genomics to improve healthcare. Led by Dr. Claire Fraser-Liggett, a preeminent genome scientist and microbiologist, IGS is located in a 10-acre BioPark on the University

of Baltimore's campus in downtown Baltimore. IGS scientists integrate genomics, bioinformatics and metagenomics into biomedical research. For more information, see www.igs.umaryland.edu.