

Bioinformatics technology provides new insight into microbial activities

Scientists may gain a new insight into the relationship between viruses and their environments thanks to a new computational technology developed by researchers at the U.S. Department of Energy's Argonne National Laboratory. This technology has already been used to identify subtle differences in the metabolic processes of microbial communities.

The ability to determine such differences may help scientists detect environmental changes at early stages and identify previously unknown pathways for treating disease.

The researchers analyzed the frequency distribution of more than 14 million microbial and viral sequences from almost 90 different ecological communities, called metagenomes. By doing so, the researchers hoped to produce a biological profile for the samples taken from diverse environments ranging from underground mines to sea and fresh water.

"Metagenomics enables the DNA from all microbes to be sequenced at once, without any culturing," said Robert Edwards, a computational biologist at Argonne and San Diego State University and one of the project's principal investigators. "Such an approach was impossible even a decade ago."

While the researchers had expected to find similar lifestyles among the viral metagenomes in every environment, they instead found that the metagenomes have distinctive metabolic profiles. Researchers may be able to use these profiles in the future to answer questions about the viral dynamics in, for example, the lungs of cystic fibrosis patients.

"Argonne has become a world leader in metagenomics," said Edwards. "The bioinformatics technology developed by Argonne researchers and their collaborators is being used by hundreds of researchers worldwide. This work demonstrates the practical basis for the multimillion-dollar effort by the National Institutes of Health to understand the benign and malign roles of microbes in health and disease."

As the use of metagenomics has become increasingly common, scientists have had to address the challenge of analyzing an enormous number of genomic sequences. To ease this process, scientists at Argonne and the Fellowship for Interpretation of Genomes (FIG) developed a system that contains all known DNA and protein sequences. Using this directory, known as SEED, biologists can identify matches between metagenomes and profiles already in the SEED database.

For this study, DNA sequences first were analyzed by using a high-throughput pipeline called the metagenomics RAST (Rapid Annotation using Subsystem Technology) server, developed by researchers from Argonne in collaboration with FIG, the University of Chicago, San Diego State University and Hope College.

"Comparing such a huge number of metagenomes is an enormous computational task," said Rick Stevens, a principal investigator in the project and associate laboratory director for Computing, Environment, and Life Sciences at Argonne. "This automated technology revolutionizes the steps needed to acquire an accurately annotated genome."

The sequences then were compared to the SEED platform by using the compute cluster at the National Microbial Pathogen Data Resource. The database allows an overview of the microbial communities and the ability to focus on one metabolic area and detect differences in the proteins being used by the microbes in each environment.

"The initial analysis took months of computer time," said Stevens. "We eventually determined that more than 1 million sequences from the microbial metagenomes and more than 500,000 from the viral metagenomes were significantly similar to functional genes within the SEED."

The results have been accepted for publication in the journal *Nature* and appear online at <http://dx.doi.org/10.1038/nature06810>.

The research was funded by the Gordon and Betty Moore Foundation Marine Microbial Initiative, the National Science Foundation, a Department of Commerce ATP grant, a National Research Initiative Competitive Grant from the U.S. Department of Agriculture Cooperative State Research, Education and Extension Service, the National Institute of Allergy and Infectious Diseases, the National Institutes of Health and the Department of Health and Human Services.

Source: Argonne National Laboratory

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