
Los Alamos creates bioinformatics tool for metagenome analysis

March 18, 2015



‘GOTTCHA’ tool could aid ID of co-infections in medical samples

LOS ALAMOS, N.M., March 18, 2015—Scientists at Los Alamos National Laboratory have developed a new method for DNA analysis of microbial communities such as those found in the ocean, the soil, and our own guts.

“Metagenomics is the study of entire microbial communities using genomics, such as when you sequence the DNA of a whole community of organisms at once,” said Patrick Chain, the lead Los Alamos scientist on the project. “The result is an enormous data set of short sequences, or ‘reads,’ that you need to sort through to try to understand which organisms are actually present, and what they may be doing. Here at Los Alamos, we

specialize in incredibly large data sets, we know how to handle them whether it's for physics, ocean or climate modeling, or for complex biological insights.

“We have developed a new tool in this rapidly expanding and evolving field of what is called ‘metagenomics’.” said Chain, “it uses nucleic acid data and looks for sections that map uniquely to a preconstructed database.”

In a paper this week in the journal *Nucleic Acids Research*, “[Accurate read-based metagenome characterization using a hierarchical suite of unique signatures](#),” the researchers present this novel method for shotgun metagenomic read classification, a method that is highly accurate, and outperforms all other most recent methods, they say.

“We believe this method will be a useful resource for analyzing metagenomic data, particularly in the area of diagnostics, where both high false-negative and false-positive rates cannot be tolerated, and where a profile of the relative abundance of certain organisms may be important,” said Chain. This method, or some version of it, is one step in the right direction toward ascertaining the presence of potential pathogens in a complex background, such as assessing medically relevant co-infections in clinical samples.

The tool, named GOTTCHA (for Genomic Origins Through Taxonomic CHallenge), makes use of a database of reference genomes that have been pre-processed to retain only unique segments of the genomes at any level of taxonomy, and then it classifies the individual metagenome sequences or “reads.” They have established a unique method to query these databases using any open access alignment software, and provide the presence and relative abundance profiles of the organisms found within a sample (community).

This is the first effort that: 1) uses a wide array of synthetic, spiked and real datasets to both train and test the utility of a read-based community profiling method; 2) importantly, provides a series of defined and realistic (in amount and quality) metagenome datasets that can be used to re-validate any current or future tools; and 3) addresses the issue of false positives which hampers most other available software. The GOTTCHA tool provides the ability to find both bacterial and viral sequences within complex samples, and makes the method flexible to database search strategies such that it can be an enduring method of community profiling.

The [software and associated databases](#), as well as training datasets used within the manuscript are accessible online.

Link to journal article: [Accurate read-based metagenome characterization using a hierarchical suite of unique signatures](#)

Los Alamos National Laboratory

www.lanl.gov

(505) 667-7000

Los Alamos, NM

Operated by Los Alamos National Security, LLC for the Department of Energy's NNSA

