

Software developed by Boston College lab delivers speed and accuracy to genome research

It took a global corps of scientists approximately \$500 million and 13 years to identify the more than 35,000 genes of the human genome. Five years later, Boston College Biologist Gabor Marth and his research team have developed software that can analyze half a million DNA sequences in 10 minutes.

The Marth laboratory's proprietary PyroBayes software is one of a new breed of computer programs able to accurately process the mountains of genome data flowing from the latest generation of gene decoding machines, which have placed a premium on computational speed and accuracy in data-crunching fields known as bioinformatics and high-throughput biology, said Marth, an associate professor of Biology.

"We're on the edge of a real technological revolution that I think will help us understand the genetic causes of diseases in humans and how genetic materials determine traits in animals," said Marth. "It is going to lead to less expensive technologies that will allow researchers to decode any individual."

PyroBayes will aid researchers involved in the 1,000 Genomes Project, which announced last month a plan to sequence the genomes of 1,000 individuals from around the world. The NIH, which helps direct the project, has awarded Marth more than \$1.3 million to develop software over the next four years.

The advances of the Marth lab were revealed in two articles published by the professor and his assistants in the February issue of Nature Methods, the premier journal of scientific research methodology.

In an article co-authored by Marth, post-doctoral researcher Chip Stewart, and graduate students Aaron Quinlan and Mike Strömberg, the group unveiled the lab's PyroBayes base caller software, which examines data from one of the latest generation of DNA decoding machines – from Roche / 454 Life Sciences – faster and with far greater accuracy than other programs for pyrosequencing, a technology that utilizes the detection of pyrophosphate for decoding the sequence of DNA, the carrier of genetic information in living organisms.

A second Nature Methods article, written in collaboration with colleagues from the Washington University School of Medicine, reported that three other computer programs developed by the Marth lab made it possible to quickly and accurately examine the whole genome of a laboratory worm and identify key differences between the sample strain and an earlier strain – a comparative process known as re-sequencing, now being applied to the genomes of humans and other organisms. This second study used another next-generation DNA sequencing platform, the Illumina/Solexa machine.

Advances are driving re-sequencing costs down, but researchers must still prove the effectiveness of the new technology by working with smaller organisms, which made the worm study critical, Marth said. "This brings us closer to a major milestone in human individual re-sequencing – the decoding of the genome of human beings in routine fashion," said Marth.

Of the few computer programs available for the new sequencing machines, the software package developed by the Marth lab is the only one capable of working with a variety of decoding machines and offers greater accuracy, allowing researchers to separate true genetic variations from data errors, said Marth. PyroBayes, a Linux-based package, is made available to fellow academic researchers at no cost.

As a member of its analysis group, the Marth lab participates in the data analysis of the 1000 Genomes Project, which was launched last month. The goal of the project is to sequence the genomes of at least 1,000 people from around the world to create the most detailed and medically useful picture to date of human genetic variation.

Ultimately, advances in bioinformatics will help push genetic science forward, shedding new light on human health and disease. Marth sees his lab's role in providing critical tools that help researchers to organize data, interpret them, and visualize genome variations.

"We are excited to develop the software that will help these super-fast, high-throughput sequencing machines to realize their potential to produce invaluable data for research," Marth said.

Source: Boston College

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.