

Accelerating genome analysis

February 22 2016

An international team of scientists, led by researchers from A*STAR's Genome Institute of Singapore (GIS) and the Bioinformatics Institute (BII), have developed SIFT 4G (SIFT for Genomes) - a software that can lead to faster genome analysis. This development was published in the scientific journal *Nature Protocols*.

Genome sequencing has been instrumental in improving knowledge of human diseases, by allowing scientists to understand their underlying biological mechanisms. It has also been critical in the global push towards precision medicine, where the genetic profile of a disease or patient could improve clinical decision-making in determining prognoses and the type of treatment prescribed for patients, paving the way for improved healthcare outcomes.

While technological advances have enabled the generation of vast amounts of data on the human body and other organisms, an issue facing the scientific community is the ability to analyse such great amounts of data well.

Based on the same principles as its predecessor, SIFT 4G can prepare predictions for genomes at a much higher speed. Once the predictions are ready, they are stored in a database and are ready for [genome analysis](#). The use of graphics processing units (GPUs) instead of central processing units (CPUs) has resulted in much shorter prediction times and the ability to construct databases for a large number of organisms. SIFT 4G already has predictions available for more than 200 organisms, thereby serving a larger research community.

While it previously took 25 days on 10 CPUs to create a database for SIFT predictions, it now takes just three days on a single GPU using SIFT 4G. By using a GPU instead of a CPU, the processing time on a protein decreased substantially from 4.2 minutes to 2.6 seconds. Once the predictions are ready, the researchers can use the database to analyse the genomes in just five minutes.

The increased efficiency in genome analysis will empower the research community in advancing science and developing technology that can benefit human health. The bioinformatics and genomic capabilities developed at BII and GIS enable both institutes to play a key role in quickly and accurately interpreting biological data to understand gene function, their interactions and the development of diseases.

Besides its obvious benefits to improving healthcare outcomes, [genome sequencing](#) has brought about significant advances in agricultural fields, as well as in [basic research](#). Sequencing different breeds or strains of the same organism can lead to an understanding of the genetic basis of its observable characteristics.

For example, over 3,000 rice genomes were sequenced to interpret the genetic diversity that underlies traits such as cold tolerance and grain quality. These types of surveys can generate a huge impact, leading some to estimate the sequencing market in agricultural and other industrial applications to be valued at more than S\$7 billion. Basic research also benefits from the rise of genome sequencing; for instance, the sequence of the *Drosophila* genome has been applied to better understand the fundamentals of evolutionary processes.

Senior corresponding author of the research, GIS' Dr Pauline Ng - who created the SIFT algorithm with her PhD advisors, Prof Steven Henikoff and Jorja Henikoff, at the Fred Hutchinson Cancer Research Center (FHCRC) more than 15 years ago - is thrilled with what SIFT 4G can

offer.

Dr Ng said, "I'm excited that SIFT 4G will make faster discovery possible for researchers studying mutations in their organism of choice. Any researcher who is studying genetic variation in an organism (whose genome has been sequenced) can now characterise their missense mutations with SIFT 4G. Its expanded functionality for many genomes will now enable researchers worldwide to use it as a resource across diverse projects."

BII Executive Director Dr Frank Eisenhaber said, "Novel algorithms and processor architectures are essential for coping with the huge amounts of data collected in biological databases, which often outpaces the advances in computer performance. This is just one of the ways in which bioinformatics and computational biology can be utilised to advance research and understanding of biological processes."

GIS Executive Director Prof Ng Huck Hui said, "Time is of the essence, especially more so when it comes to the research fields. We have to be able to adapt quickly in order to keep up with the constantly evolving healthcare and biomedical landscape; SIFT 4G is a powerful tool for us to do so. It will accelerate the time taken for genome analysis and that can only benefit the [research community](#) and the public at large."

Provided by Biomedical Sciences Institutes (BMSI)

Citation: Accelerating genome analysis (2016, February 22) retrieved 6 May 2024 from <https://phys.org/news/2016-02-genome-analysis.html>

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.