

naturenews

Published online 17 November 2008 | Nature | doi:10.1038/news.2008.1235

News

The search for genome 'dark matter' moves closer

The multi-million dollar 1000 Genomes project is set to be finished in a year.

Brendan Maher

An almost complete catalogue of human genetic variation could be available by the end of 2009, thanks to a massive genome sequencing project that includes academic and industrial partners around the world.

Announcing completion of the pilot phase of the 1000 Genomes project, the project's co-chair David Altshuler said last week that it has already successfully catalogued 3.8 trillion bases of sequence — approximately a thousand times the number found in a single human genome. Speaking at the annual meeting for the American Society of Human Genetics in Philadelphia, Pennsylvania, Altshuler predicated that the data should be deposited in GenBank, the US National Institutes of Health database for genetic sequences, within the next two months. But that's only around a tenth of the total amount of data the project aims to rack up by the time it has finished.

"One thing we're good at is having ambitious goals," said Alan Gutmacher, interim director of the US National Human Genome Research Institute (NHGRI), one of the organizations funding the project — the final cost of which could reach US\$50 million.

Variations on a theme

The sequencing initiative is a direct descendant of efforts to sequence the human genome that started in the 1990s. Soon after the announcement of the genome's first draft in 2000, a drive to map all common genetic variations between people began. This effort, known as the International HapMap Project, produced a reference guide of hundreds of millions of single-letter differences among people, which could be used to quickly compare genomes without directly sequencing them.

Versions of this map have been used in genome-wide association studies during recent years to find genetic variants that are associated with common diseases, and could point to ways of predicting or even treating those diseases.

But the HapMap details only the most common genetic variants — those that occur in more than 1 in 20 people — and contains little information on rare single nucleotide polymorphisms (SNPs) or on so-called structural variants, such as large block deletions or duplications of gene sequence, which seem to be important in some diseases.

This is where 1000 Genomes comes in, says Altshuler, who is also a professor of genetics and medicine at Harvard Medical School in Boston, Massachusetts. "Next-generation sequencing makes it possible to examine a different part of the allele spectrum," he said. By fully sequencing some 1,200 people from various ethnic groups around the world and looking at parent-child trios alongside data from other sequencing projects, 1000 Genomes hopes to capture rarer genetic variations — namely, those that occur in fewer than 5% and more than 1% of people and some that occur at even smaller frequencies. The work comparing parents and their children could help scientists to get better estimates of the individual mutation rate from generation to generation. "We'll get an unparalleled view of human genetic variation," says Richard Durbin of the Wellcome Trust Sanger Institute in Cambridge, UK, and the other co-chair of 1000 Genomes.

Getting personal

The 1000 Genomes project will contain no detailed demographic or medical information about the people being sequenced so that data can be shared without the need for complicated consent procedures. Meanwhile, another major sequencing effort — the Personal Genome Project, headed by George Church of Harvard Medical School — aims to provide full sequences with medical and personal information for up to 100,000 subjects. Although its goals are less centred on uncovering variation, the Personal Genome Project, together with 1000 Genomes and several other projects looking at genetic variation, will significantly increase the amount of available DNA data for analysis.

The pilot phase of the 1000 Genomes project, which sequenced more than 180 individuals, has identified 4 million SNPs, 22% of which



The same but different. The 1000 Genomes project aims to catalogue human genetic variation.

Punchstock

seem to be previously undiscovered. These data will be released in December and January, and quarterly releases are expected throughout 2009 up to the completion of the data-collection phase at the end of next year.

But the near-petabyte levels of data to be collected (by some estimates) pose significant challenges for storage and accessibility, as well as the analysis and usability of the data, says David Haussler, who leads the genome bioinformatics group at the University of California, Santa Cruz. In a nod to this need, the NHGRI announced this month that it will make up to \$14 million available during the next two years for data handling and analysis for the project.

In search of 'dark matter'

Once a better catalogue of variation is complete, it could be used to power the next generation of genome-wide association studies to understand disease, potentially filling in some of the so-called missing heritability (see '[Personal genomes: The case of the missing heritability](#)') – the genetic markers for traits or diseases that current association studies have been unable to find. These missing genetic variants have been called the genome's 'dark matter'. But some believe that the rare variants that 1000 Genomes aims to turn up may not provide useful information about disease.

"1000 Genomes will be hugely useful for growing the technology to generate and analyse sequence data," says David Goldstein of Duke University in Durham, North Carolina, adding "But in terms of a catalogue of the variants most important to human biology and disease, it's less clear how important it will be." Goldstein advocates sequencing people with extreme presentations of disease to understand more about common disease pathways.

Altshuler disagrees with Goldstein but is also cautious. "None of us imagines that we will explain 100% of disease heritability when this is finished, nor will there be drugs in the clinic immediately."

Geneticists, however, are excited about the prospect of exploring DNA's dark matter in a year's time. "I don't think it's going to be dark matter for too long," said David Valle, a clinical geneticist at Johns Hopkins University School of Medicine in Baltimore, Maryland. "When the light comes, I think we're going to find some interesting biology."

Comments

Reader comments are usually moderated after posting. If you find something offensive or inappropriate, you can speed this process by clicking 'Report this comment' (or, if that doesn't work for you, email redesign@nature.com). For more controversial topics, we reserve the right to moderate before comments are published.

Add your own comment

You can be as critical or controversial as you like, but please don't get personal or offensive, and do keep it brief. Remember this is for feedback and discussion - not for publishing papers, press releases or advertisements, for example. If you ramble on in an annoying way too often, we may remove your posting privileges.

You need to be registered with Nature to leave a comment. Please log in or register as a new user. You will be re-directed back to this page.

[Log in / register](#)

Nature ISSN 0028-0836 | EISSN 1476-4687

© 2008 Nature Publishing Group – partner of [AGORA](#), [HINARI](#), [CrossRef](#) and [COUNTER](#)

ADVERTISEMENT



PREDICTING
VOLCANIC
ERUPTIONS AT
THE MOUTH OF
THE FURNACE

Laureate
Andrew McGonigle

[LEARN MORE](#)

2008 ROLEX AWARDS