

Scientists sequence genomes from Africa, Europe, Asia

October 27 2010



This undated handout illustration shows the DNA double helix. US researchers for the first time have developed a technique to sequence and compare the genomes of 179 people representing the continents of Africa, Europe and Asia, a study said Wednesday.

Small genetic differences between individuals help explain why some people have a higher risk than others for developing illnesses such as diabetes or cancer. Today in the journal *Nature*, the 1000 Genomes Project, an international public-private consortium, published the most comprehensive map of these genetic differences, called variations, estimated to contain approximately 95 percent of the genetic variation of any person on Earth.

Researchers produced the map using next-generation DNA sequencing technologies to systematically characterize human genetic variation in 180 people in three pilot studies. Moreover, the full scale-up from the

pilots is already under way, with data already collected from more than 1,000 people.

"The pilot studies of the 1000 Genomes Project laid a critical foundation for studying human genetic variation," said Richard Durbin, Ph.D., of the Wellcome Trust Sanger Institute and co-chair of the consortium.

"These proof-of-principle studies are enabling consortium scientists to create a comprehensive, publicly available map of genetic variation that will ultimately collect sequence from 2,500 people from multiple populations worldwide and underpin future genetics research."

Genetic variation between people refers to differences in the order of the chemical units — called bases — that make up DNA in the human genome. These differences can be as small as a single base being replaced by a different one — which is called a single nucleotide polymorphism (abbreviated SNP) — or is as large as whole sections of a chromosome being duplicated or relocated to another place in the genome. Some of these variations are common in the population and some are rare. By comparing many individuals to one another and by comparing one population to other populations, researchers can create a map of all types of genetic variation.

The 1000 Genomes Project's aim is to provide a comprehensive public resource that supports researchers aiming to study all types of genetic variation that might cause human disease. The project's approach goes beyond previous efforts in capturing and integrating data on all types of variation, and by studying samples from numerous human populations with informed consent allowing free data release without restriction on use. Already, these data have been used in studies of the genetic basis for disease.

"By making data from the project freely available to the research community, it is already impacting research for both rare and common

diseases," said David Altshuler, M.D., Ph.D., Deputy Director of the Broad Institute of Harvard and MIT, and a co-chair of the project.

"Biotech companies have developed genotyping products to test common variants from the project for a role in disease. Every published study using next-generation sequencing to find rare disease mutations, and those in cancer, used project data to filter out variants that might obscure their results."

The project has studied populations with European, West African and East Asian ancestry. Using the newest technologies for sequencing DNA, the project's nine centers sequenced the whole genome of 179 people and the protein-coding genes of 697 people. Each region was sequenced several times, so that more than 4.5 terabases (4.5 million million base letters) of DNA sequence were collected. A consortium involving academic centers on multiple continents and technology companies that developed and sell the sequencing equipment carried out the work.

To process these data required many technical and computational innovations, including standardized ways to organize, store, analyze and share DNA sequencing data. Launched in 2008, the 1000 Genomes Project started with three pilot projects to develop, evaluate and compare strategies for producing a catalogue of genetic variations. Funded through numerous mechanisms by foundations and national governments, the 1000 Genome Project will cost some \$120 million over five years, ending in 2012.

When the work began, sequencing was very expensive, so the project began with two approaches aimed at increasing efficiency: One strategy, called "low-pass", combines partial data from many people; the second, only focused on the part of the genome that encodes protein-coding genes. By comparing these strategies to "gold standard" data produced at great completeness and accuracy, the project was able to show that both the alternative approaches work well and have complementary strengths.

Researchers will use both strategies in the full-scale project because, although sequencing costs have decreased, it is still relatively expensive.

"We have shown for the first time that a new approach to sequencing — low coverage of many samples — works efficiently and well," said Gil McVean, Ph.D., Professor of Statistical Genetics at the University of Oxford. "This proof of principle is now being applied not only in the 1000 Genomes Project, but in disease research, as well."

The resulting map of human genetic variation includes about 15 million SNPs, 1 million short insertion/deletion changes, and more than 20,000 structural variations. Many of the genetic variants had previously been identified, but more than half were new. The project's database contains more than 95 percent of the currently measurable variants found in any individual, and continuing work will eventually identify more than 99 percent of human variants.

Richard Gibbs, Ph.D., director of the Human Genome Sequencing Center at the Baylor College of Medicine (one of the project's sequencing centers) said, "What really excites me about this project is the focus on identifying variants in the protein-coding genes that have functional consequences. These will be extremely useful for studies of disease and evolution."

The improved map produced some surprises. For example, the researchers discovered that on average, each person carries between 250 and 300 genetic changes that would cause a gene to stop working normally, and that each person also carried between 50 and 100 genetic variations that had previously been associated with an inherited disease. No human carries a perfect set of genes. Fortunately, because each person carries at least two copies of every gene, individuals likely remain healthy, even while carrying these defective genes, if the second copy works normally.

In addition to looking at variants that are shared between many people, the researchers also investigated in detail the genomes of six people: two mother-father-daughter nuclear families. By finding new variants present in the daughter but not the parents, the team was able to observe the precise rate of mutations in humans, showing that each person has approximately 60 new mutations that are not in either parent.

With the completion of the pilot phase, the 1000 Genomes Project has moved into full-scale studies in which 2,500 samples from 27 populations will be studied over the next two years. Data from the pilot studies and the full-scale project are freely available on the project web site, www.1000genomes.org.

Researchers studying specific illnesses, such as heart disease or cancer, use maps of genetic variation to help them identify genetic changes that may contribute to the illnesses. Over the last five years, the first generation of such studies (called genome-wide association studies or GWAS) have been based on an earlier map of genetic variation called the HapMap. Built using older technology, HapMap lacks the completeness and detail of the 1000 Genomes Project.

"The 1000 Genomes Project map fills in the gaps between the HapMap landmarks, helping researchers identify all candidate genes in a region associated with a disease," said Lisa Brooks, Ph.D., program director for genetic variation at the National Human Genome Research Institute, a part of the National Institutes of Health. "Once a disease-associated region of the genome is identified, experimental studies must be done to identify which variants, genes, and regulatory elements cause the increased disease risk. With the new map, researchers can just look up all the candidate genes and almost all of the variants in the database, saving them many steps in finding the causes."

Provided by MIT

Citation: Scientists sequence genomes from Africa, Europe, Asia (2010, October 27) retrieved 10 April 2024 from

<https://medicalxpress.com/news/2010-10-scientists-sequence-genomes-africa-europe.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.