

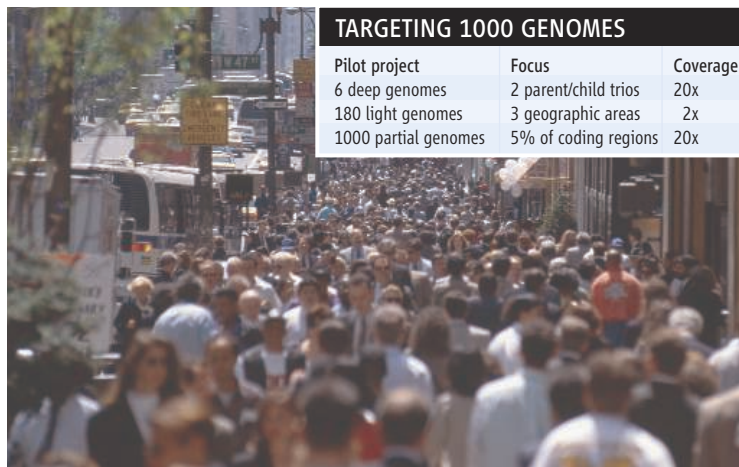
DNA SEQUENCING

A Plan to Capture Human Diversity in 1000 Genomes

It's sign of how fast horizons are changing in biology: Researchers who only a few years ago were being asked to justify the cost of sequencing a single human genome are now breezily offering to sequence 1000. And they say they can do it in a flash. Over the next 3 years, an international team plans to create a massive new genome catalog that will serve as "a gold-standard reference set for analysis of human variation," says Richard Durbin of the Sanger Institute in Hinxton, U.K., who proposed the project just last year.

The 1000 Genomes Project, as it's called, will delve much deeper than the sequencing of celebrity genomes, three of which were completed last year. It will help fill out the list of new genetic markers for common diseases that came out in 2007, says Francis Collins, director of the U.S. National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. At the same time, new technologies will be put to the test, and researchers will work out how to handle a growing deluge of data. Such practical advances will be needed a few years from now when sequencing entire genomes will be routine, notes population geneticist Kenneth Weiss of Pennsylvania State University in State College, who is not part of the project. "This seems overall like a next logical step," he says.

The search for disease genes took off last year, building on the first human genome reference sequence in 2003 and the subsequent HapMap. The latter describes how blocks of DNA tagged by common variants, called single-nucleotide polymorphisms (SNPs), vary in different populations. These SNPs have turned up more than 100 new DNA markers associated with common illnesses such as diabetes and heart disease (*Science*, 21 December 2007, p. 1842). But the HapMap includes only the most common markers, those present in at least 5% of the population.



More is better. Researchers aim to acquire DNA data from 1000 individuals.

TARGETING 1000 GENOMES

Pilot project	Focus	Coverage
6 deep genomes	2 parent/child trios	20x
180 light genomes	3 geographic areas	2x
1000 partial genomes	5% of coding regions	20x

To find rarer SNPs that occur at 1% frequency, genome leaders say, they need to sequence about 1000 genomes. According to a plan hammered out by about three dozen experts last year, the project will take advantage of new technologies that have slashed the cost of sequencing. The work will be done by the three U.S. sequencing centers funded by NHGRI, the Sanger Institute, and the Beijing Genomics Institute (BGI) in Shenzhen, China.

Because the technologies are so new, the consortium will start with three pilot projects. One will exhaustively sequence the entire genome of six individuals: two adults and both sets of their parents. DNA in these six genomes will be analyzed repeatedly up to 20 times to ensure almost complete coverage. A second project will sequence 180 individual genomes at light (2x) coverage, leaving gaps. The third project will be to fully sequence (20x coverage) the protein-coding regions of 1000 genes (5% of the total) in about 1000 genomes. The samples, all anonymous and with no clinical information, will mainly be drawn from those collected for the HapMap, which includes people of European, Asian, and African descent.

The pilots should take about a year and will put the new technologies to a "very vigorous test," Collins says. After that, the consortium will decide what coverage to use to sequence the entire set of 1000 genomes. Most of the project's \$30 million to \$50 mil-

lion price tag will be paid from the existing sequencing budgets of institutes, organizers say.

The new catalog could help disease gene hunters in several ways. It may allow researchers simply to hunt through an index for a SNP in a particular location that alters a gene product rather than run a time-consuming sequencing project, Collins says. The project will also catalog genes that are sometimes lost or duplicated; such copy-number variants can cause disease. By compiling rarer variants, it should also help resolve a

debate about the relative contribution of these mutations to disease risks. "There's no question it's going to be a tremendous resource," says Yale University's Judy Cho, who has used the HapMap to find a new gene for Crohn's disease.

China is also launching its own human genomes project. BGI Shenzhen this month announced that it is seeking 99 volunteers who will help pay to have their genomes sequenced as part of a study of diversity (*Science*, 26 October 2007, p. 553). The 3-year effort, called the Yanhuang Project after the Yan and Huang tribes that are believed to be ancestors of modern Chinese, will overlap with the 1000 Genomes Project. With proper consent, some volunteers' genomes will be sequenced for both efforts, says Wang Jun, director of BGI Hangzhou.

In a parallel effort, J. Craig Venter of the J. Craig Venter Institute in Rockville, Maryland, says his team will sequence up to 10 individuals this year and publish the data along with medical information. Venter—who dismisses the 1000 Genomes Project as "more survey work" because not all genomes will be sequenced to great depth—has even bolder plans. He says he aims for "complete diploid genome sequencing" of 10,000 human genomes in the next decade. Still, he says, "it's great that there's such an expansion of things."

—JOCELYN KAISER

With reporting by Hao Xin.