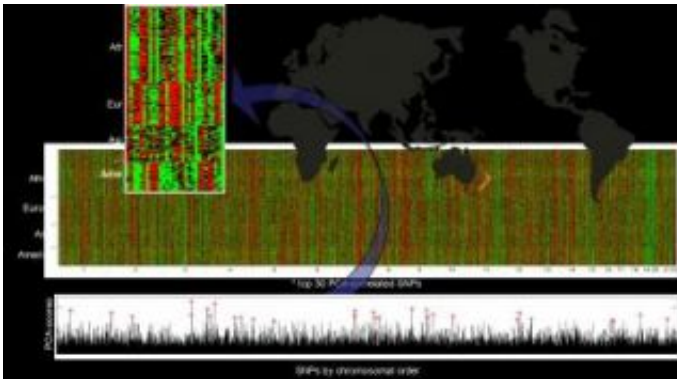# Computer program traces ancestry using anonymous DNA samples

September 21 2007

Plot of genetic markers for 255 individuals from four continental regions. Red and green represents identical genotypes. Black represents genotypic variations. Notice the distinct patterns formed in the four continental blocks, highlighting the genetic similarities between people of the same ancestry. Credit: Peristera Paschou

A group of computer scientists, mathematicians, and biologists from around the world have developed a computer algorithm that can help trace the genetic ancestry of thousands of individuals in minutes, without any prior knowledge of their background. The team's findings will be published in the September 2007 edition of the journal *PLoS Genetics*.

Unlike previous computer programs of its kind that require prior knowledge of an individual's ancestry and background, this new algorithm looks for specific DNA markers known as single nucleotide

polymorphisms, or SNPs (pronounced snips), and needs nothing more than a DNA sample in the form of a simple cheek swab. The researchers used genetic data from previous studies to perform and confirm their research, including the new HapMap database, which is working to uncover and map variations in the human genome.

"Now that we have found that the program works well, we hope to implement it on a much larger scale, using hundreds of thousands of SNPs and thousands of individuals," said Petros Drineas, the senior author of the study and assistant professor of computer science at Rensselaer Polytechnic Institute. "The program will be a valuable tool for understanding our genetic ancestry and targeting drugs and other medical treatments because it might be possible that these can affect people of different ancestry in very different ways."

Understanding our unique genetic makeup is a crucial step to unraveling the genetic basis for complex diseases, according to the paper. Although the human genome is 99 percent the same from human to human, it is that 1 percent that can have a major impact on our response to diseases, viruses, medications, and toxins. If researchers can uncover the minute genetic details that set each of us apart, biomedical research and treatments can be better customized for each individual, Drineas said.

This program will help people understand their unique backgrounds and aid historians and anthropologists in their study of where different populations originated and how humans became such a hugely diverse, global society.

Their program was more than 99 percent accurate and correctly identified the ancestry of hundreds of individuals. This included people from genetically similar populations (such as Chinese and Japanese) and complex genetic populations like Puerto Ricans who can come from a variety of backgrounds including Native American, European, and

African.

"When we compared our findings to the existing datasets, only one individual was incorrectly identified and his background was almost equally close between Chinese and Japanese," Drineas said.

In addition to Drineas, the algorithm was developed by scientists from California, Puerto Rico, and Greece. The researchers involved include lead author Peristera Paschou from the Democritus University of Thrace in Greece; Elad Ziv, Esteban G. Burchard, and Shweta Choudhry from the University of California, San Francisco; William Rodriguez-Cintron from the University of Puerto Rico School of Medicine in San Juan; and Michael W. Mahoney from Yahoo! Research in California.

Source: Rensselaer Polytechnic Institute