

Scientists Sequence 64 Human Genomes as New Reference for Genetic Diversity

(https://www.genengnews.com/wp-content/uploads/2019/05/March_2021/Apr30_2014_32859465_DNAMultiColored_FeatArticleEpigeneticsPlot.jpg)

An international team of collaborating scientists has published details of their work to sequence 64 full human genomes, representing 25 different human populations from across the globe. The researchers claim the new reference dataset represents the full range of genetic variant types. It could facilitate population-specific studies on the genetic predisposition to different human diseases, and may inform on the development of novel strategies for personalized medicines targeted to individual genetic makeup. Importantly, the team says, each of the genomes was assembled without guidance from the first human genome, and so should better capture genetic differences from different human populations.

The study, reported in *Science*, was led by scientists from the European Molecular Biology Laboratory Heidelberg (EMBL), the Heinrich Heine University Düsseldorf (HHU), the Jackson Laboratory for Genomic Medicine (JAX), and the University of Washington in Seattle (UW). “Here, we present a resource consisting of phased genome assemblies, corresponding to 70 haplotypes (64 unrelated and 6 children) from a diverse panel of human

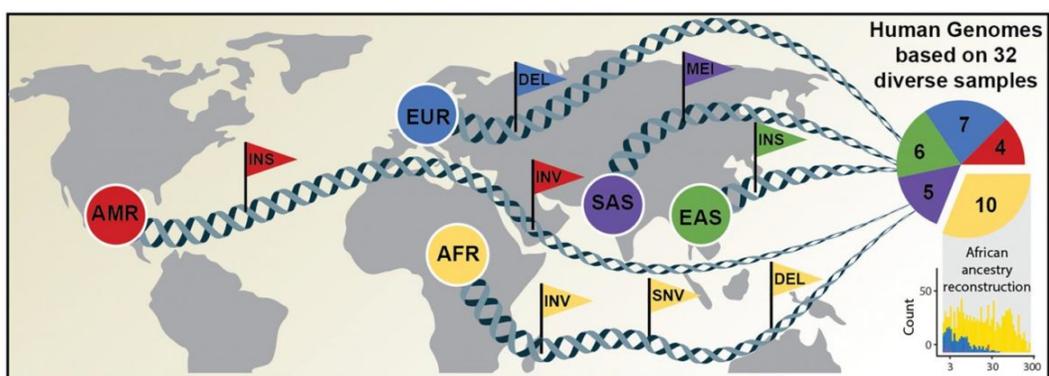
genomes,” the authors wrote in their published paper. “We have generated a diversity panel of phased long-read human genome assemblies that has significantly improved SV [structural variant] discovery and will serve as the basis to construct new population-specific references. ...The work provides fundamental new insights into the structure, variation, and mutation of the human genome providing a framework for more systematic analyses of thousands of human genomes going forward.”

Co-first author, Peter Ebert, PhD, from the Institute of Medical Biometry and Bioinformatics at HHU, stated, “With these new reference data, genetic differences can be studied with unprecedented accuracy against the background of global genetic variation, which facilitates the biomedical evaluation of genetic variants carried by an individual.” The team’s paper is titled, “Haplotype-resolved diverse human genomes and integrated analysis of structural variation (<http://dx.doi.org/10.1126/science.abf7117>).”

Twenty years ago last month, the International Human Genome Sequencing Consortium announced the first draft of the human genome reference sequence. The Human Genome Project, as it was called, involved more than 1,000 scientists from 40 countries, and took more than 11 years. However, the initial reference genome did not represent a single individual, but instead was a composite of humans, and this could not accurately capture the complexity of human genetic variation.

Over the intervening 20 years, and building on the initial results, scientists have conducted several sequencing projects to identify and catalog genetic differences between an individual and the reference genome. Those differences usually focused on small single base changes and missed larger genetic alterations. Current technologies are now beginning to detect and characterize larger differences—called structural variants—such as insertions of new genetic material. Structural variants are more likely than smaller genetic differences to interfere with gene function. The distribution of genetic variants can also differ substantially between population groups as a result of spontaneous and continuously occurring changes in the genetic material. If such a mutation is passed on over many generations, it can become a genetic variant specific to that population.

The newly reported, more comprehensive reference dataset was generated using a combination of advanced sequencing and mapping technologies. “Previous large-scale efforts have largely be inferential and biased when it comes to the detection of SVs,” said study co-author Scott Devine, PhD, associate professor of medicine at University of Maryland School of Medicine (UMSOM) and faculty member of IGS, who added, “We’ve entered a new era in genomics where whole human genomes can be sequenced with exciting new technologies that provide more substantial and accurate reads of the DNA bases. This is allowing researchers to study areas of the genome that previously were not accessible but are relevant to human traits and diseases.”



Comprehensive discovery of genetic variation based on analysis of human genomes of diverse ancestry. [David Porubsky, University of Washington]

The study builds on a method published by the researchers last year in *Nature Biotechnology* (<https://www.nature.com/articles/s41587-020-0719-5>), to accurately reconstruct the two components of a person’s genome—one inherited from a person’s father, one from a person’s mother. When assembling a person’s genome, this method eliminates the potential biases that could result from comparisons with an imperfect reference genome. So, unlike previous population surveys of structural variation, the Phased Assembly Variant (PAV) caller can discover genetic variants through direct comparison between the two sequence-assembled haplotypes and the human reference genome. “Here, we develop a method to discover all forms of genetic variation (PAV) directly by comparison of assembled human genomes,” they wrote.

The reference data generated by the new study also provide an important basis for including the full spectrum of genetic variants in genome-wide association studies. The aim is to estimate the individual risk of developing

certain diseases such as cancer and to understand the underlying molecular mechanisms. This, in turn, might then be used as a basis for more targeted therapies and preventative medicine.

The Institute of Genome Science (IGS) Genome Resource Center (GRC) was one of three sequencing centers, along with Jackson Labs and the University of Washington, which generated the data, using new sequencing technology that was developed recently by Pacific Biosciences. The GRC was one of only five early access centers that was asked to test the new platform.

Devine helped to lead the sequencing efforts for the study, and also led the sub-group researchers and co-authors who discovered the presence of “mobile elements” (i.e., pieces of DNA that can move around and get inserted into other areas of the genome). “We characterized 130 of the most active mobile element source elements,” Devine and colleagues wrote. Other members of the Institute for Genome Sciences (IGS) at UMSOM are among the 65 co-authors. Luke Tallon, PhD, scientific director of the Genomic Resource Center, worked with Devine to generate one of the first human genome sequences on the Pacific Biosciences platform that was contributed to this study. Nelson Chuang, a graduate student in Devine’s lab also contributed to the project.

“The landmark new research demonstrates a giant step forward in our understanding of the underpinnings of genetically-driven health conditions,” said E. Albert Reece, MD, PhD, executive vice president for medical affairs, University of Maryland, Baltimore, and the John Z. and Akiko K. Bowers distinguished professor and dean, UMSOM. “This advance will hopefully fuel future studies aimed at understanding the impact of human genome variation on human diseases.”
