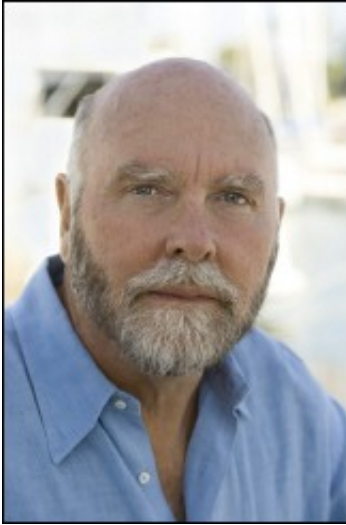


# First individual genome sequence published

September 4 2007

---



This file photo shows Dr. J. Craig Venter on his research sail boat, the Sorcerer II, off the coast of San Diego California. The first individual genome ever sequenced -- a complete DNA blueprint of celebrity scientist Venter -- has revealed genetic variation among humans far richer than previously imagined.

Independent sequence and assembly of the six billion base pairs from the genome of one person ushers in the era of individualized genomics.

Researchers at the J. Craig Venter Institute (JCVI), along with collaborators at The Hospital for Sick Children in Toronto and the University of California San Diego (UCSD), have published a genome sequence of an individual, Craig Venter, that covers both sets of chromosomes that were inherited from each parent.

Two other versions of the human genome currently exist—one published in 2001 by J. Craig Venter, Ph.D., and colleagues at Celera Genomics, and another at the same time by a consortium of government-funded researchers. These genomes were not of any single individual, but, rather, were a melding of DNA from various people. In the case of Celera, it was a consensus assembly from five individuals, while the government-funded version was a haploid genome based on sequencing from a limited number of individuals. Both versions greatly underestimated human genetic diversity.

This new genome, known as the “HuRef” version, represents the first time a true diploid genome from one individual—Dr. Venter—has been published. The research is available in the latest issue of the open-access journal *PLoS Biology*.

Researchers at the JCVI have been sequencing and analyzing this version of Dr. Venter’s genome since 2003. Building on reanalyzed data from Dr. Venter’s genome that constituted 60% of the previously published Celera genome, the team had the goal of constructing a true reference human genome based on one individual. Using whole genome shotgun sequencing and highly accurate long reads from Sanger dideoxy automated DNA sequencing, the team produced additional data making the final 32 million sequences.

From the combined data set of more than 20 billion base pairs, the researchers were able to assemble the human genome with an overall length of 2.810 billion base pairs. The genome was covered 7.5 times, ensuring that each set of contributing chromosomes was covered over 3.2 times for greater than 96% coverage of the two parental genomes. The team at JCVI compared and contrasted the new HuRef diploid genome sequence to earlier versions of published human genomes and found that the HuRef version improved upon both these early versions by providing more and correctly oriented base pairs.

Since the HuRef genome is diploid, each of the parental chromosomes could be directly compared to each other. One of the most surprising and important findings from this research was the high degree of genetic variation that was found between two chromosomes within a single individual.

“Each time we peer into the human genome, we uncover more valuable insight into our intricate biology,” said Dr. Venter. “With this publication, we have shown that human-to-human variation is more than seven-fold greater than earlier estimates, proving that we are in fact very unique individuals at the genetic level.” He added, “It is clear, however, that we are still at the earliest stages of discovery about ourselves, and only with continued sequencing of more individual genomes will we be able to garner a full understanding of how our genes influence our lives.”

Within the human genome, there are different kinds of DNA variants. The most studied type is single nucleotide polymorphisms, or SNPs. These have long been thought to be the most prevalent and perhaps the most important type of variant implicated in human traits and disease susceptibility. However, in this analysis of Dr. Venter’s genome, the team found a surprising number of other important variants. A total of 4.1 million variants covering 12.3 million base pairs of DNA were uncovered with more than 1.2 million new variants discovered.

Of the 4.1 million variations between chromosome sets, 3.2 million were SNPs, while nearly one million were other kinds of variants, such as insertion/deletions (“indels”), copy number variants, block substitutions, and segmental duplications. While the SNPs outnumbered the non-SNP types of variants, the non-SNP variants involved a larger portion of the genome. This suggests that human-to-human variation is much greater than previously thought. The researchers suggest that much more research needs to be done on these non-SNP variants to better understand their role in individual genomics.

According to Sam Levy, Ph.D., lead author and senior scientist at JCVI, “The ability to use unbiased, high throughput sequencing methods, coupled with advance computational analytic methods, enables us to characterize more comprehensively the wide variety of individual genetic variation. This offers us an unprecedented opportunity to study the prevalence and impact of these DNA variants on traits and diseases in human populations.”

Another important feature that is made possible by having an individual, diploid genome is the ability to begin to do better and more informed haplotype assemblies. Haplotypes are groups of linked variants. Through the government-sponsored HapMap project, many common haplotypes have been identified; however, these are based on averages of large ethnogeographic populations rather than individuals. Having individual haplotypes would enable researchers to understand and find more rare or individual variants that would explain and help predict diseases in that particular person—a truly personalized, individualized genomics paradigm. In the HuRef analysis, the team used the 4.1 million variant set and new algorithms to build haplotype assemblies that, when compared to the HapMap project, represented longer and more complete linkages. The JCVI researchers expect this number to improve significantly as additional sequence coverage is added to HuRef using a variety of new sequencing technologies.

Long-range haplotype linkages will enable much more complete analysis of human variation and the genetic association with complex human traits, behaviors, and diseases. In the near future, the scientists believe that it will be possible to know from which parent various traits were inherited. Already in this analysis, the JCVI team has found more than 300 disease genes and 4,000 genes overall that exhibit different protein forms. This will be an important area for further study and analysis to determine how these altered proteins affect Dr. Venter’s health status.

Citation: Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, et al. (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5(10): e254. doi:10.1371/journal.pbio.0050254.

Source: Public Library of Science

Citation: First individual genome sequence published (2007, September 4) retrieved 27 April 2024 from <https://phys.org/news/2007-09-individual-genome-sequence-published.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.