Researchers look to CNVs to better understand human ethnic diversity

August 7 2015, by Bob Yirka

A depiction of the double helical structure of DNA. Its four coding units (A, T, C, G) are color-coded in pink, orange, purple and yellow. Credit: NHGRI

(Phys.org)—A large team of international researchers has undertaken a study to better understand the genetic mechanism behind ethnic diversity. In their paper published in the journal *Science*, the team describes their study, their results and their plans for the future.

In studying the human genome, scientists have looked at genomic features called single-nucleotide variants (SNVs)—the study of them formed the basis of the 1,000 Genomes Project—but there is another feature that the researchers with this new effort suggest, can offer us information about our genome that we cannot get any other way. Copy-number variants (CNVs) have information about deletions and duplications, which can offer clues about the history of our species, and because of that, allow for tracking ethnic diversity. But CNVs are harder to study, the team notes, which is why they have not been looked at as much. In this new effort the team took on the challenge, conducting both SNV and CNV sequencing on 236 people that included 125 ethnically diverse populations around the world.

The team reports that as part of their study they found that the median size of CNV base pairs was 7,396, which meant the group had to probe up to ten times more deeply than was done with the 1,000 Genomes Project. Doing so allowed them to reconstruct parts of the ancestral history of the human genome—at 200,000 years ago, for example, they were able to see the differences between humans at the time and chimps and orangutans. It also allowed them to see that humans had approximately 40 million more base pairs than we have today.

The researchers also report that while they found that there were fewer numbers of CNV base pairs, compared to SNVs, more differentiation between groups could be seen, which they claim suggests, that they are largely responsible for ethnic diversity. Duplication, they noted, appeared to be responsible for the greatest amount of change—up to four times as much as deletions, with the biggest changes occurring in

groups that migrated out of Africa.

The team is contemplating expanding the study to look at comparing CNV loci among various people alive today, which they say, could offer new clues about disease risks among different ethnic groups.

ABSTRACT
In order to explore the diversity and selective signatures of duplication and deletion human copy number variants (CNVs), we sequenced 236 individuals from 125 distinct human populations. We observed that duplications exhibit fundamentally different population genetic and selective signatures than deletions and are more likely to be stratified between human populations. Through reconstruction of the ancestral human genome, we identify megabases of DNA lost in different human lineages and pinpoint large duplications that introgressed from the extinct Denisova lineage now found at high frequency exclusively in Oceanic populations. We find that the proportion of CNV base pairs to single nucleotide variant base pairs is greater among non-Africans than it is among African populations, but we conclude that this difference is likely due to unique aspects of non-African population history as opposed to differences in CNV load.