



Understanding the extent, distribution and age of human protein-coding genetic variants across diverse populations allows fascinating insights into human population dynamics and the resultant evolutionary forces. Cataloguing and dating such variation will also allow us to understand the origin of the seemingly endless list of potential disease variants and to prioritize among them for further investigation. A recent study describes the sequencing of 15,336 genes in 4,298 individuals of European American and 2,217 individuals of African American ancestry, providing insights into a recent human population expansion and the associated evolution of disease variants.

As a part of the US National Institutes of Health Heart, Lung and Blood Institute (NHLBI)-sponsored Exome Sequencing Project (ESP), Fu *et al.* identified 1,146,401 autosomal single-nucleotide variants (SNVs) in the sequenced exomes. They dated these using a simulation approach that modelled the data to different coalescence scenarios. The age of the SNVs was consistent with a modified 'out-of-Africa' model in which accelerated population growth began 5,115 years ago with the per-generation growth rate being higher in European Americans than in African Americans. The identification of an excess of rare variants in the data supported the occurrence of this population expansion. Furthermore, the authors estimate that 73% of all protein-coding variants occurred within the past 5,000 years. SNVs that were more than 50,000 years old were more often associated with African American samples, probably as a result of stronger genetic drift in European populations that is associated with the migration out of Africa.

To identify putative deleterious variants, Fu *et al.* used four functional prediction algorithms and two conservation-based methods. They found that ~86% of deleterious variants arose within the past 5,000 years (91.2% and 77% for European Americans and African Americans, respectively) and that the fraction of putative deleterious variants diminished with the age of the variants. The authors then analysed the distribution of these putative deleterious mutations in genes that cause Mendelian disorders, essential genes and genes that are associated with complex disease. In European American populations, the percentage of deleterious genes in these categories was increased relative to genes classified as 'other'. However, this was not the case in African American samples. Further simulations revealed that this difference was probably due to the population bottleneck associated with the out-of-Africa expansion, resulting in weaker purifying selection. The distribution of the putative deleterious variants in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways showed that the age of mutations varied across functional pathways in both population samples. For example, in metabolic pathways, variants were older than in other pathways and, interestingly, in disease pathways, variants were newer than in other pathways.

The study here provides interesting insights into the recent expansion of disease-causing variants. Furthermore, it could provide a framework on which new methods for prioritizing disease-causing variants could be based.

Hannah Stower

ORIGINAL RESEARCH PAPER Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 28 Nov 2012 (doi:10.1038/nature11690)