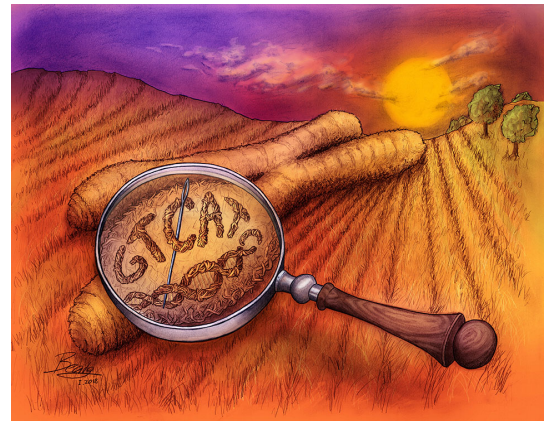


February 16, 2018 | By Ioana Patringenaru

New algorithm can pinpoint mutations in large sections of the human genome

A team of scientists has developed an algorithm that can accurately pinpoint, in large regions of the human genome, mutations favored by natural selection. The finding provides deeper insight into how evolution works, and ultimately could lead to better treatments for genetic disorders. For example, adaptation to chronic hypoxia at high altitude can suggest targets for cardiovascular and other ischemic diseases.



The sequenced genome of a single individual yields about half a terabyte of data of information—that’s about as much information as you’ll find on 106 DVDs. A population sample of size 1000 individuals contains 1000 times as much information. So to examine such a massive amount of data, researchers turned to computational techniques.

“Computer science and data science are playing a significant role to better understand the code of life and uncover the hidden patterns in our genome,” said Ali Akbari, the paper’s first author and a Ph.D. student in electrical and computer engineering at the University of California San Diego. “We are analyzing massively large sets of human genomic data to ultimately improve our understanding of genetic basis of diseases.”

Researchers detail the algorithm, dubbed iSAFE, in the Feb. 19 issue of *Nature Methods*.

Many existing genomic analysis approaches can detect which regions of the human genome are evolving under selection pressure. Often, these regions are large, covering millions of base-pairs and do not shed light on the specific mutations that are responding to the selection pressure. iSAFE doesn’t need to know the function of the genomic region it is analyzing or any

demographic information for the human population it belongs to. Instead, the researchers used population genetic signals imprinted in the genomes of the sampled individuals and machine learning techniques to reliably identify the mutation favored by selection.

In natural selection, neighboring mutations ‘hitchhike’ with the mutation that is under positive selection, leading to a loss of genetic diversity near the favored mutation. iSAFE exploits signals in the neighboring sequences, the so-called “shoulder regions” to pinpoint the favored mutation.

“Finding the favored mutation among tens of thousands of other, hitchhiking, mutations was like a needle in a haystack problem,” said Akbari, who works in the research group of computer science professor Vineet Bafna at the Jacobs School of Engineering at UC San Diego.

To test the algorithm, researchers ran iSAFE on regions of the genome that are home to known favored mutations. The algorithm ranked the correct mutation as the top one out of more than 21,000 possibilities in 69 percent of cases, as opposed to state of the art methods, which only did this in 10 percent of cases.

The algorithm also identified a host of previously unknown mutations, including five that involve genes related to pigmentation. In these cases, iSAFE identified identical mutations in multiple non-African populations. This suggests an early response to the onset of selection as humans migrated out of Africa.

The research was supported in part by the National Science Foundation and the National Institutes of Health.

Identifying the favored mutation in a positive selective sweep

Corresponding authors: Ali Akbari, Department of Electrical and Computer Engineering and Vineet Bafna, Department of Computer Science and Engineering, UC San Diego.

Co-authors: Arya Iranmehr and Siavash Mirarab, Department of Electrical and Computer Engineering, UC San Diego; Mehrdad Bakhtiari, Department of Computer Science and Engineering, UC San Diego; Joseph J. Vitti and Pardis Sabeti, Department of Organismic and Evolutionary Biology, Harvard and Broad Institute of MIT and Harvard.

DOI: 10.1038/NMETH.4606

MEDIA CONTACT

Ioana Patringenaru, 858-822-0899, ipatrin@ucsd.edu

UC San Diego's [Studio Ten 300](#) offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.