

New mathematical model to add rigor to studies of disease genetics and evolution

March 19 2007

USC College computational biologist Peter Calabrese has developed a new model to simulate the evolution of so-called recombination hotspots in the genome.

Published March 5 in the early online edition of the *Proceedings of the National Academy of Sciences*, the mathematical model and its associated software bring much-needed rigor to evolutionary investigations of how natural selection acts on individual genes, said Calabrese, a research assistant professor of biological sciences.

And, they may also aid the search for disease-associated genes within the human genome.

The new tools "are more rigorous and less time-consuming than previous, simpler models," Calabrese said.

Recent interest in genetic recombination hotspots has been fueled partly by the promise of genome association studies, which aim to locate the chromosomal regions responsible for genetic diseases. Analyzing such studies to understand the inheritance of genes associated with disease requires an understanding of genetic recombination at a very fine scale.

Genes are packaged in larger structures called chromosomes. Humans have 23 pairs of matching chromosomes, one inherited from each parent. In almost every cell in the body, the maternal and paternal chromosomes stay separated in these pairs. But the sex cells (sperm and

egg) each carry only one copy of each chromosome, a mix of the two inherited chromosomes produced by genetic recombination.

During the creation of new sperm or egg cells, paired chromosomes line up and exchange stretches of DNA before dividing into four cells, each with its own singular and completely new chromosome, which will be passed on to offspring.

This genetic re-arrangement and re-shuffling is a major source of genetic diversity, and so is considered the primary benefit of sexual reproduction. It's the biological process that makes each individual (save identical twins) unique, even from close relatives such as siblings.

To the surprise of many, recent research indicates that most recombination occurs in small regions of the genome called hotspots. As scientists have explored details of this process, it's become clear that the majority — approximately 80 percent — of recombination occurs at these narrow bands of activity, only 1,000 to 2,000 DNA bases wide. Hotspots make up only 10 to 20 percent of the human genome, which runs about 3 billion DNA bases long. Rates of recombination at a hotspot may be as much as hundreds to thousands of times that of the surrounding gene sequence. Little is known about hotspot origins or how they work.

Scientists have identified a small number of human recombination hotspots over the last few years. An important advance came in 2005, when an Oxford University team estimated the location of approximately 25,000 potential hotspots on the human genome. In doing so, they assumed the locations of hotspots would not differ greatly between individuals.

However, comparisons of large numbers of human genomic data (including the International HapMap Project, which created rough maps

of the genomes of hundreds of people from all around the globe) have revealed a much more complex picture of hotspots across the human population. And, work by geneticist Norman Arnheim, a USC Distinguished Professor and the Ester Dornsife Chair in Biological Sciences in the College, and others shows that hotspots, like genes themselves, do vary across the population.

Arnheim, one of Calabrese's collaborators, runs one of a handful of laboratories in the world that uses the painstaking but powerful method of sperm typing to study genetic recombination. He and others have shown that some hotspots are heterogeneous — not everyone has the same the hotspots at the same locations.

Calabrese's model and software take these differences, as well as the chance that the rate of recombination might not be constant over time, into account, where older models did not.

His work helps explain a number of puzzles confronting those who study hotspots.

The first is that while chimpanzees, our closest primate relative, share 99 percent of their genetic code with humans, studies have revealed almost no overlap in hotspots in their genome.

"The chimp-human comparison really was a surprise," Calabrese said. "Even with a very similar DNA sequence, the chimps' hotspots appear completely independent of humans."

Calabrese's model fits with and helps to explain this finding. Since the last common ancestor of chimpanzees and humans lived 6 to 7 million years ago, the model predicts that enough time has passed for humans to evolve a distinct set of hotspots.

The model also fits with human evidence. Data from the HapMap project, for example, shows that African-Americans and Asian-Americans have differences in the locations and frequency of genomic hotspots, findings backed up by other studies in a number of ethnic groups. But they also share many hotspots.

Only about 100,000 years have passed since the last major human migration out of Africa, Calabrese writes in his paper, which his model predicts is not enough time for geographically separated populations to have evolved completely unique sets of hotspots.

To Calabrese, one of the most exciting applications of his model is how it might inform the discussion of the "hotspot paradox."

A confusing aspect of hotspot origination, the paradox considers how hotspot locations are tagged. Previously, researchers identified at least one short sequence of DNA bases, called a DNA motif, associated with about 10 percent of hotspots in humans. The paradox arises from the idea that if the "tag" or motif lies too close to the hotspot, the tag is likely to be lost within a few generations due to the high rate of DNA breakage at and near the site.

"Obviously, something other than the sequence itself probably codes for these hotspots," he said. Another hypothesis is that the DNA motif is located a good distance away from the hotspot.

Calabrese's model also considers other types of signaling mechanisms hotspots might employ, such as epigenetic (for example, molecular tags attached to the DNA, but not the DNA sequence itself) or multiple, interacting DNA motifs.

The simulation software allows scientists to compare DNA sequences to find hotspot patterns in the population, which may be important to

understanding disease or evolution.

His coalescent simulation computer program, available free for download at Calabrese's Web site, showed that existing simulation software can reliably detect the most common (present in 50 percent or more of individuals) hotspots in large human genetic datasets, but probably miss the majority of rarer (present in less than 10 percent of individuals) hotspot sites.

Elucidating the location of and how recombination hotspots work is critical to building fundamental understanding of the biological mechanisms that promote genetic variation, Calabrese notes. Indirectly, the knowledge also may inform the work of scientists designing better, faster ways to search for genes thought to play a role in human disease, he said.

Source: University of Southern California

Citation: New mathematical model to add rigor to studies of disease genetics and evolution (2007, March 19) retrieved 19 April 2024 from <https://phys.org/news/2007-03-mathematical-rigor-disease-genetics-evolution.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.