Homozygosity Runs in our DNA

Durgadas Kasbekar
INSA Senior Scientist
Centre for DNA Fingerprinting and Diagnostics, Hyderabad

Our genome comprises 23 DNA molecules. Each is the backbone of one of our 23 chromosomes (numbered 1 to 23). And our cells have two copies of each chromosome. One copy inherited from the mother, and the other from the father. Maternally- and paternally-derived pairs are called homologues. If the DNA in the 23 paternally-derived homologues were stretched out and laid down end-to-end, they would span about 2 meters. But they are packaged to fit into the head of a sperm cell, whose longest dimension is about 6 microns (millionth of a meter).

The two members of a homologue pair are unambiguously similar to each other. That is, their DNA molecules have pretty much the same length and sequence of chemical bases. But they are not identical. They differ at sites in which one or the other DNA molecule has suffered a mutation in its ancestry. Mutations are any changes in the DNA base sequence, and most occur randomly over time. Any sizeable stretch of identical DNA sequence retained by both homologues is called a run of homozygosity or ROH.

Consanguinous marriage is a union between related individuals. Children from consanguinous unions have an increased probability of inheriting the same long chromosome stretch from both parents. Thus, consanguinity produces long ROH. Numerous shorter ROH are seen in the genome of individuals from populations that practice endogamy, which is the custom of marriage only within the community, clan, or tribe. Smaller numbers of short ROH are also present in the genome of children of unrelated parents, since the parents, like any two people, must at least share distant ancestors. Humans and chimpanzees shared a common ancestor ~5-7 million years ago, and their genomes differ by about 4%. A virtual “offspring” from human and chimpanzee parents also would show many still shorter ROH.

A study of 1043 individuals from Sub-Saharan Africa, Middle East, Europe, Central and South Asia, East Asia, Oceania, and Central and South America

(https://doi.org/10.1371/journal.pone.0013996) compared the number of short and long ROH in their genome with the sum length of the genome in ROH. The fraction of the genome in short ROH correlated with the overland distance from Addis Ababa. But not the fraction in long ROH. Long ROH, reflecting recent inbreeding, were prominent in populations practicing consanguinity or endogamy. This result can be explained by the origin of humans in Africa, followed by their dispersal in small groups into continental Eurasia, and still later secondary dispersals from Eurasia to Oceania and America. Since emigrants contain only a subset of the source population's mutations the migrations created genetic bottlenecks. At one extreme, one-third of the genome of a Karitiana individual from the Brazilian Amazon was in ROH, whereas at the other extreme, a Mandinka from Senegal had 35-fold less ROH.

The Mandinka and other ethnic groups were captured, enslaved, and shipped from West Africa to the Americas in the 16th to 19th century. While today their descendants enrich the cultural canvas of Brazil, the Caribbean, and the United States, the history of slavery is a painful and shameful reminder of how, despite our lofty pieties and philosophies, we easily lose our humanity. The Karitiana are a small indigenous population of fewer than 500 isolated individuals eking a living in the rain forest. In the late 1970s scientists took their blood samples to create cell lines. The cells were then commercially distributed to other scientists to perform studies like the one described above. We like to think the blood was obtained after proper "informed consent". But the allegations of trickery and coercion are not implausible (https://www.nytimes.com/2007/06/20/world/americas/20blood.html). That these backstories bookend values of a parameter which establishes shared ancestry is an irony missed by many readers of the scientific paper.

Let us now turn to the 39 DNA molecules of the dog genome, possibly our DNA's best friends. We, dogs, and wolves, shared a common ancestor 100 million years ago. About 20,000 to 40,000 years ago dogs began to become domesticated alongside humans, and to diverge from wolves. The specialized breeding of dogs took off only about 200 years ago. Both domestication and breed formation created severe genetic bottlenecks. The genome fraction in ROH is greater in dogs than humans and in breed dogs than street dogs. Only 27 genes were not found in any ROH among 4342 dogs examined (https://www.pnas.org/content/118/16/e2019116118/tab-article-info). Some founder dogs might

have carried a homologue with a strongly deleterious mutation that when homozygous (i.e., inherited from both parents) rendered the offspring inviable.

Thus, deleterious mutations can lead genome regions to be underrepresented in the ROH. This drives conversion of long ROH into shorter ROH flanking the mutation. Over time, long ROH become shorter. An ROH can be reduced down to say a stretch of only 200 bases of DNA, as in a virtual progeny of a human and fish. Even so, it is most easily explained by a distant common ancestor. Any alternative hypothesis is either more convoluted or just wrong.