

Can most Americans be identified by a relative's DNA? Maybe soon

October 12 2018



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

The remarkable technique used to identify the suspected "Golden State Killer" four decades after his crimes—genetic genealogy—could be used to identify half of all Americans from relatives' DNA samples, a new study says.

And only a few years from now, the process could be used to track nearly all Americans of European descent by making DNA matches with distant relatives, the authors of the study predict.

The research, published Thursday in the US journal *Science*, could have wide-ranging privacy implications—if someone uses a consumer website to trace his ancestry, should that information be used to identify his kin, possibly in a criminal case?

"We are on our way to get to the point that virtually anyone will have a third cousin in those databases," said Yaniv Erlich, the chief science officer at the MyHeritage website, and senior author of the study.

"I predict it will happen within two to three years."

A person and his or her third cousin have the same great-great-grandparents. With a second cousin, one shares great-grandparents.

The closer you are with a relative, the more similar your genetic make-up is.

Even in the case of third cousins, the human genome—or the information encoded in a person's DNA—is very much alike.

Genealogy and police work

When police find a DNA sample that does not match anyone in their database, a criminal investigation can come to a dead end.

In California, police had been at that point for decades in the case of the so-called Golden State Killer, who is blamed for 12 murders and more than 50 rapes dating back to the mid-1970s.

Then they uploaded his DNA sample to a free website called GEDmatch, which allows users to post DNA test results in text format.

The site then generates a list of people with similar genomes, ranked from the closest to the most distant—with names and email addresses.

In the Golden State case, investigators hit the jackpot—the suspected killer's third cousins popped up as a match.

Police rebuilt the family trees as far back as the 1800s... before wading through the hundreds of descendants to try to find their suspect.

By eliminating possible relatives by sex, age or residence, they landed on Joseph James DeAngelo, whose DNA they discreetly obtained from a car door handle and his trash.

That sample matched one left at the scene of a 1980 murder. DeAngelo is now behind bars awaiting trial.

Since that breakthrough, police departments across the country are using these techniques to try to resolve their cold cases.

Thirteen people have been arrested in five months, according to Parabon NanoLabs, a company that analyzed 200 mystery samples.

According to the company's director of bioinformatics, Ellen McRae Greytak, 60 percent of those samples had "matches" on GEDMatch that were worth pursuing.

Parabon's researchers work assiduously using publically available data (genealogy websites, Facebook accounts, LinkedIn profiles, obituaries, etc) to rebuild family trees and identify possible suspects.

Beyond the 13 cases that led to arrests, "we have several other ones where we've given them a lead of a single individual," Greytak told AFP.

Legal limbo

For the study published Thursday, researchers analyzed the DNA data of the 1.28 million people in the MyHeritage database.

MyHeritage is one of several websites that offers DNA analysis from saliva samples for a fee. Others include AncestryDNA and 23andMe.

So far, most of those who paid to use these services (MyHeritage charges \$79-99) are white.

Researchers discovered that 60 percent of Americans of European descent had a "match" with a third cousin or someone even more closely related.

That means that with samples from only two percent of the total US population, all could be identified.

Unlike GEDMatch (1.1 million files), other sites like Ancestry (10 million) and 23andMe (5 million) are not open to public searches.

One day, police could order the sites to open up their databases. So far, Ancestry and 23andMe told AFP they have yet to receive an injunction.

But the threat of such action in the future, or the illegal use of another person's DNA, worries privacy advocates.

"It really re-emphasizes the need for people to fully understand what is going to happen to their data if they upload it on these sites," said Benjamin Berkman, a bioethics researcher at the National Institutes of Health.

Natalie Ram, a professor at the University of Baltimore School of Law, says she hopes that the new research will help raise awareness about the legal void that allowed police to use GEDMatch to find DeAngelo.

Ram says genetic data should be constitutionally protected from illegal search and seizure much like a person's email or telephone data.

"It will have to be worked out in court," she predicted.

Erlich, the author of the study, says he plans to get ahead of a potential crisis, to make sure that genetic genealogy does not become the focus of a leak scandal on par with the data breaches suffered by large companies such as Facebook.

He proposes that each [sample](#) be sealed with some sort of encrypted signature that would prevent unauthorized usage.

"I am concerned we will have some moment of reckoning—'Oh, we should have done something five years ago'," he says.

More information: Y. Erlich et al., "Identity inference of genomic data using long-range familial searches," *Science* (2018).

science.sciencemag.org/lookup/.../1126/science.aau4832

© 2018 AFP

Citation: Can most Americans be identified by a relative's DNA? Maybe soon (2018, October

12) retrieved 28 June 2024 from <https://phys.org/news/2018-10-americans-relative-dna.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.