

A new strategy to speed up cold case investigations

September 16 2022, by Dave Gilson



Investigators trying to ID an unknown person may spend months combing through massive family trees. Credit: Cory Hall

For nearly 37 years, she was known as Buckskin Girl—a young, anonymous murder victim found outside Dayton, Ohio, wearing a deer-hide poncho. Then, in April 2018, police announced that the mystery of her identity had been solved. Her name was Marcia L. King, and she had

been identified by linking a snippet of her DNA to one of her cousins.

It was one of the first high-profile cases in which this investigative method had been used to identify an unclaimed body. Two weeks after King's name was revealed, police in California announced that they'd used similar techniques to track down the Golden State Killer. Suddenly, the combination of genetic sampling, genealogical research, and old-fashioned gumshoeing was hailed as a revolutionary breakthrough that would crack hundreds of cold cases.

Since then, forensic genetic genealogy has cleared more than 400 cases in the U.S. Yet this detective work is complex and time-consuming. While King was IDed after just a few hours of sleuthing, most cases take much longer. On average, they take over a year to solve successfully. Many are left unfinished: Law enforcement agencies may run out of funding before a person can be identified and investigators may give up if they hit too many dead ends.

A more systematic approach would help, says Lawrence Wein, a professor of operations, information, and technology at Stanford Graduate School of Business. With Mine Su Ertürk, Ph.D., he has unveiled a method for solving cold cases with more speed and success. In a new paper in the *Journal of Forensic Sciences*, they present the first detailed mathematical analysis of the forensic genetic genealogy process and outline a way for investigators to optimize searches for unknown victims or criminal suspects.

To develop their search method, Wein and Ertürk teamed up with the [DNA Doe Project](#), a California nonprofit that has solved more than 65 cases of unidentified remains, including the King case. It provided the researchers with data from 17 cases, including 8 that were unsolved at the time. "That's quite similar to the historical average of cases they've solved," Wein says. "So there's no reason to suspect that these cases are

much harder or much easier than randomly selected cases."

Using that real-world data, Wein and Ertürk looked at how forensic genetic genealogy searches are commonly done and then tested their method, which aims to maximize the probability of finding a solution in the shortest time possible. "It turns out to be much quicker," Wein says of the new approach—nearly 10 times quicker. "If they're only solving a small number of cases using the current method, and we can get them to solve them 10 times faster, then they could solve a lot more cases."

A forest of family trees

A typical genetic genealogy investigation begins with a DNA sample from a "target" such as an unidentified body or a murder suspect. It's uploaded into a DNA database such as GEDmatch or FamilyTreeDNA, which generates a list of "matches"—people who share pieces of the target's genome. A search may turn up hundreds of these matches, usually distant cousins whose shared ancestors may have died more than a century ago. The cases Wein and Ertürk analyzed had between 200 and 5,000 matches.

That's just the start: Drawing a line from these far-flung relatives to the target requires building a [family tree](#) that includes as many family members as possible. Here, too, the scale of the problem is daunting. "These are huge trees," Wein says. "It's really hard to visually lay out anything bigger than a couple dozen people." As the tree expands, the odds of identifying the target improve—but the length of the search also increases.

Next, the relevant people in the tree have to be identified. This requires scouring [public records](#), genealogy sites, and social media—time-intensive legwork that combines intuition and skill. "There's a whole art to it," Wein says. "Using marriage documents and death documents and

birth documents and Facebook and all kinds of different records to try to figure out who people are and who their ancestors and offspring are."

It's not immediately apparent which matches will provide the best path to the target. Investigators' strategies to follow these leads tend to be decentralized, Wein says. "You have a team of people doing this and they will each decide to take a match to investigate, and then they'll go off on their own to try to build a family tree backward in time from each match. They're not thinking about the big picture holistically."

By stepping back and assessing the entire problem, Wein and Ertürk provide a roadmap for genetic genealogists seeking the most efficient path to an unidentified target. "Basically, we're telling them, "Given where you are in the search right now, this is what you should do next," Wein says.

Solving equations and crimes

Explaining the difference between the new search method and the standard, or "benchmark," method is complicated, but Wein boils it down to this: "The benchmark method looks for common ancestors between different matches. What you really want to find is the most recent common ancestor between a match and the unknown target, and that's a slightly different problem." The most recent common ancestor of first cousins, for example, is a grandparent; second cousins share a great-grandparent, and so on.

After identifying a list of possible most recent common ancestors, Wein and Ertürk's method "aggressively" fills out the family tree with their descendants, even if there's only a slight chance that the target's ancestor is on the list.

This leap is accomplished by using probability theory to track the

search's progress. "We do this by describing the reconstructed family tree as a collection of probabilities that represent how likely each person on our tree is to be a correct ancestor of the target," Ertürk explains. "Then, looking at these probabilities, you can tell which parts of the tree you should explore more."

This approach proves effective even with smaller family trees, which means faster solve times. After running hundreds of simulated searches, Wein and Ertürk conclude that their method can solve a case with a 7,500-person family tree around 94% of the time. The standard method's success rate in those cases is around 4%.

Wein hopes these findings will help the DNA Doe Project and other investigators refine their approach and crack more cases. He notes that his analysis doesn't account for some of the "tricks" genetic investigators use to narrow their searches, such as focusing on [family members](#) who lived in a particular location. "In no way is our algorithm meant to substitute for genealogists," he says. "But if they're really stuck, it will give them some ideas that may be non-obvious."

Wein and Ertürk have applied mathematics to tackle investigative challenges before. Last year, Ertürk cowrote a paper with Stanford GSB professor Kuang Xu that outlined a method of genetic searching that balances efficiency and privacy concerns. Wein has researched better ways to scan fingerprints, process sexual assault kits, and trace bullets.

He sees forensic genetic genealogy as another crime-solving tool that can be improved so it might live up to its promise. "It's an interesting field that combines probability and statistics and optimization and sometimes game theory," he says. "That's how I, from a mathematical point of view, stayed attracted to these problems."

More information: Mine Su Ertürk et al, Analysis of the genealogy

process in forensic genetic genealogy, *Journal of Forensic Sciences*
(2022). [DOI: 10.1111/1556-4029.15127](https://doi.org/10.1111/1556-4029.15127)

Provided by Stanford University

Citation: A new strategy to speed up cold case investigations (2022, September 16) retrieved 22
June 2024 from <https://phys.org/news/2022-09-strategy-cold-case.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.