

# Record-breaking DNA comparisons drive fast forensics

June 27 2019, by Anne MCGovern

---



With the IdPrism platform, algorithms encode genetic markers as bits to allow for fast DNA comparisons in forensics. “When Hollywood finds out about this, they’re going to change their scripts,” says Darrell Ricke of the Bioengineering Systems and Technologies Group. “The capabilities are so different from what’s out there.” Credit: Thomas Sheehan

Forensic investigators arrive at the scene of a crime to search for clues. There are no known suspects, and every second that passes means more time for the trail to run cold. A DNA sample is discovered, collected, and then sent to a nearby forensics laboratory. There, it is sequenced and fed into a program that compares its genetic contents to DNA profiles stored in the FBI's [National DNA Index System](#) (NDIS)—a database containing profiles of 18 million people who have passed through the criminal justice system. The hope is that the crime scene sample will match a profile from the database, pointing the way to a suspect. The sample can also be used for kinship analysis through which the sample is linked to blood relatives, as was done last April to catch the infamous [Golden State Killer](#).

DNA forensics is a powerful tool, yet it presents a computational scaling problem when it is improved and expanded for complex samples (those containing DNA from more than one individual) and kinship analysis. Consider the volume of data that the FBI must handle for the nation. "If you think of all the [police stations](#) across the country, all operating each week, it's a lot of data to keep track of and organize," says Darrell Ricke from the Bioengineering Systems and Technologies Group. To put this into perspective, if each state compares 2,000 crime scene samples weekly, that's 100,000 samples to compare against 18 million profiles per week.

Ricke is part of a team at the laboratory that developed an integrated web-based platform called IdPrism that provides expanded comparison capabilities without compromising speed or functionality. IdPrism allows identification of more than 10 individuals in a complex DNA sample, along with extended kinship results. At its heart are two algorithms that Ricke developed, FastID and TachysSTR, which encode [genetic markers](#) as bits (0 or 1) and operate quickly and smoothly. These algorithms recently won a 2018 R&D 100 Award, which is given annually by R&D Magazine to the 100 most significant inventions of the year.

These markers are two types of variations in DNA called short tandem repeats (STR) and single nucleotide polymorphisms (SNP). They are considered to be a kind of DNA fingerprint that can be used to identify individuals as well as their relatives. Each person has a unique combination of SNP or STR variations—one person's combination presents in a specific pattern, while another person's presents in a different pattern. When analysts run a crime-scene DNA sample against a profile in the NDIS database, finding a matching combination of these STRs shows a high chance that the DNA belongs to the same person.

The FBI currently uses software algorithms that must pass through a complex set of calculations to reveal if a sample matches a profile. Ricke's algorithms assign a bit value to normal (0) or rare (1) versions of SNPs, or a bit for each different STR marker. The normal label indicates that the SNP or STR is common in many people and is thus not a unique marker that can be used to identify an individual. With this digital DNA encoding for both identity comparisons and complex mixtures, analysis can be done with just three hardware bit instructions: exclusive OR, logical AND, and population count.

An exclusive OR instruction allows for a comparison of whether two DNA profiles are the same or different. For the forensic comparisons, this instruction will output a 0 when an SNP or STR in a sample matches that in a profile, and it will output a 1 when they don't match. This technique works well when the crime scene sample contains DNA from only one individual, but if there are more contributors, a matching result could be hidden among mismatches from the other people in the same sample. This issue is addressed by adding a logical AND with the database profile to the results of the exclusive OR. This step, in a sense, gets rid of the mismatch noise to reveal whether the database profile has matched against an individual in the sample. The final step is population count, which sums up all of the 1s. In the end, a match is represented by mostly 0s and a mismatch will have a high number of 1s.

Using these three hardware bit instructions, the FastID algorithm can compare 5,000 SNPs in a crime scene DNA sample against 20 million reference profiles in under 12 seconds. Alternative methods would take hours to do so on this scale. Similarly, TachysSTR can compare STRs in 1 million samples in 1.8 seconds, whereas current algorithms take 10 minutes to do the same.

The results are displayed inside the IdPrism system in which investigators can run, view, query, and store their DNA comparison data. In addition to being fast and convenient, the system has improved the accuracy of forensics by including a panel of 2,650 SNP markers that are used for complex sample and kinship analysis.

Last November, the system was transitioned to users outside of the laboratory. "Although getting IdPrism to a transition-ready product was challenging, it is awesome to think that our technology is being used," says Philip Fremont-Smith, who is also from the Bioengineering Systems and Technologies Group and was involved in the bioinformatics side of the project.

"When Hollywood finds out about this, they're going to change their scripts," Ricke says. "The capabilities are so different from what's out there."

*This story is republished courtesy of MIT News ([web.mit.edu/newsoffice/](http://web.mit.edu/newsoffice/)), a popular site that covers news about MIT research, innovation and teaching.*

Provided by Massachusetts Institute of Technology

Citation: Record-breaking DNA comparisons drive fast forensics (2019, June 27) retrieved 24 June 2024 from <https://phys.org/news/2019-06-record-breaking-dna-comparisons-fast-forensics.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.