

# NIST builds statistical foundation for next-generation forensic DNA profiling

July 23 2018



NIST scientists have published statistical data needed for forensic DNA profiling based on a technology called Next Generation Sequencing. To do that, they sequenced forensic DNA markers for a sample population. The letters A, G, T, and C represent the building blocks of DNA. Credit: K. Irvine, NIST

DNA is often considered the most reliable form of forensic evidence, and this reputation is based on the way DNA experts use statistics. When they compare the DNA left at a crime scene with the DNA of a suspect, experts generate statistics that describe how closely those DNA samples

match. A jury can then take those match statistics into account when deciding guilt or innocence.

These match statistics are reliable because they're based on rigorous scientific research. However, that research only applies to DNA fingerprints, also called DNA profiles, that have been generated using current technology. Now, scientists at the National Institute of Standards and Technology (NIST) have laid the statistical foundation for calculating match statistics when using Next Generation Sequencing, or NGS, which produces DNA profiles that can be more useful in solving some crimes. This research, which was jointly funded by NIST and the FBI, was published in *Forensic Science International: Genetics*.

"If you're working criminal cases, you need to be able to generate match statistics," said Katherine Gettings, the NIST biologist who led the study. "The data we've published will make it possible for labs that use NGS to generate those statistics."

## **How to Create a DNA Profile**

To generate a DNA profile, forensic labs analyze sections of DNA, called genetic markers, where the genetic code repeats itself, like a word typed over and over again. Those sections are called short tandem repeats, or STRs, and the number of repeats at each marker varies from person to person. The analyst doesn't actually read the genetic sequence inside those markers, but just counts the number of repeats at each one. That yields a series of numbers that, like a long social security number, can be used to identify a person.

STR-based profiling was developed in the 1990s, when genetic sequencing was hugely expensive. Today, NGS makes sequencing cost-effective for biomedical research and other applications. NGS can also be used to create forensic DNA profiles that, unlike traditional STR

profiles, include the actual genetic sequence inside the markers. That provides a lot more data.

That extra data might not be needed because in most cases, STR-based profiles contain more than enough information to reliably identify a suspect. However, if the evidence contains only a minute amount of DNA, or if the DNA has been exposed to the elements and has begun to break down, then the analyst might only get a partial profile, which may not be enough to identify a suspect. In those cases, the extra data in an NGS-based profile might help solve the case.

In addition, evidence that contains a mixture of DNA from several people can be difficult to interpret. The extra data in NGS-based profiles can help in those cases as well.

## **Calculating Match Statistics**

DNA analysts are able to calculate match statistics for STR-based profiles because scientists have measured how frequently different versions of the markers occur in the population. With those frequencies, you can calculate the chances of randomly encountering a particular DNA profile, just as you can calculate the chances of picking all the right numbers in a lottery.

NIST measured those STR gene frequencies years ago using a library of DNA samples from 1,036 individuals. To calculate gene frequencies for NGS-based profiles, Gettings and her co-authors cracked open the freezer that contained the original samples, which were anonymized and donated by people who consented to their DNA being used for research. The scientists generated NGS-based profiles for them by sequencing 27 markers—the core set of 20 included in most DNA profiles in the U.S. plus seven others. They then calculated the frequencies for the various genetic sequences found at each marker.

It might be surprising that scientists can estimate gene frequencies from such a small library of samples. However, the NIST team was measuring frequencies not for the full profiles, but for the individual markers. Since they sequenced 27 markers, with each marker occurring twice per sample, the number of markers tested wasn't 1,036, but more than 55,000.

Although NIST has now published the data needed to generate match statistics for NGS-based profiles, other hurdles must still be cleared before the new technology sees widespread use in forensics. For instance, labs will have to develop ways to manage the greater amounts of data produced by NGS. They will also have to implement operating procedures and quality controls for the new technology. Still, while much work remains, said Peter Vallone, the research chemist who leads NIST's forensic genetics research, "We're laying the foundation for the future."

**More information:** Katherine Butler Gettings et al, U.S. Population Sequence Data for 27 Autosomal STR Loci, *Forensic Science International: Genetics* (2018). [DOI: 10.1016/j.fsigen.2018.07.013](https://doi.org/10.1016/j.fsigen.2018.07.013)

Provided by National Institute of Standards and Technology

Citation: NIST builds statistical foundation for next-generation forensic DNA profiling (2018, July 23) retrieved 6 May 2024 from <https://phys.org/news/2018-07-nist-statistical-foundation-next-generation-forensic.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.