New DNA database at Rutgers-Camden to strengthen forensic science

January 23 2018, by Jeanne Leong

Forensic DNA evidence is a valuable tool in criminal investigations to link a suspect to the scene of a crime, but the process to make that determination is not so simple since the genetic material found at a crime scene often comes from more than one person.

That task may become somewhat less challenging, thanks to a new database at Rutgers University-Camden that can help to bring more reliability to the interpretation of complex DNA evidence. This innovative new resource was developed by a research team led by Rutgers University-Camden professors Catherine Grgicak and Desmond Lun, and Ken Duffy of the University of Ireland at Maynooth.

"Right now, there's no standardization of tests," says Grgicak, the Henry Rutgers Chair in chemistry at Rutgers-Camden. "There's accreditation of crime labs, but that's different from having standards set out for labs to meet some critical threshold of a match statistic."

In analyzing DNA mixtures, scientists will often find partial matches, so part of the determination of whether a suspect contributed to an item of evidence depends on interpretations by forensic scientists.

The Project Research Openness for Validation with Empirical Data (PROVEDIt) database will help reduce the risk of misinterpreting the profile. The database is online at https://lftdi.camden.rutgers.edu/provedit.

The team of researchers spent more than six years developing computational algorithms that sorted through possible DNA signal combinations in a piece of evidence, taking into account their prevalence in the general population to determine the likelihood that the genetic material came from one, two, three, four, or five people.

Information from the PROVEDIt database, the housed at Rutgers-Camden, could be used to test software systems and interpretation protocols, and be used as a benchmark for future developments in DNA analysis.

The PROVEDIt database, which consists of approximately 25,000 samples, is accessible to anyone for free.

"We wanted to provide these data to the community so that they could test their own probabilistic systems," says Grgicak. "Other academicians or other researchers might develop their own systems by which to interpret these very complex types of samples."

The website's files contain data that can be used to develop new or compare existing interpretation or analysis strategies.

Grgicak says forensic laboratories could use the database for validating or testing new or existing forensic DNA interpretation protocols. Researchers requiring data to test newly developed methodologies, technologies, ideas, developments, hypotheses, or prototypes can use the database to advance their own work.

Lun, a computer science professor at Rutgers-Camden, led the way in developing the software systems, doing the number crunching to determine the likely number of contributors in a DNA sample, and calculating statistics to determine the likelihood that a person contributed to a sample or not.

"The approach that we took to develop these methods is that we thought that it is very important that they be empirically driven," says Lun. "That they can be used on real experimental data in order both to train or calibrate these methods and validate them."

Grgicak's and Lun's research to produce the database, titled "A Large-Scale Dataset of Single and Mixed-Source Short Tandem Repeat Profiles to Inform Human Identification Strategies: PROVEDIt," is published in the journal *Forensic Science International: Genetics*.

The database was mentioned in 2016 in a report by President Barak Obama's President's Council of Advisors on Science and Technology (PCAST), an advisory group of the nation's leading scientists and engineers who directly advise the president and make policy recommendations in science, technology, and innovation.

More information: Lauren E. Alfonse et al. A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, *Forensic Science International: Genetics* (2017). DOI: 10.1016/j.fsigen.2017.10.006

Provided by Rutgers University