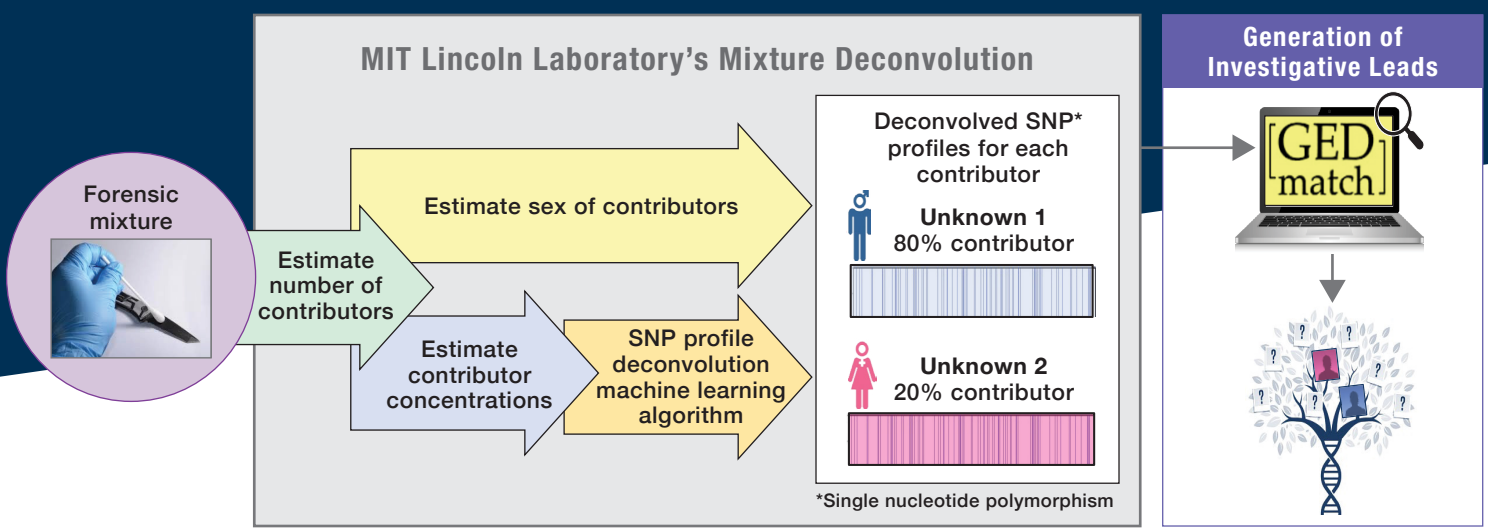
MIT LINCOLN LABORATORY

INNOVATION HIGHLIGHT

Technology in Support of National Security

WWW.LL.MIT.EDU

Finding Familial Connections in a Mixed DNA Sample



In the workflow for identifying familial matches in a two-person DNA mixture, machine learning algorithms are applied to (1) estimate the number of contributors in a DNA mixture, (2) estimate the percent contributions of each contributor, (3) estimate the sex of each contributor, (4) unravel (deconvolve) different DNA profiles in the mixture to isolate the two contributors, and then (5) match those two DNA profiles to potential family members.

Lincoln Laboratory collaborated with partner Verogen to develop a process for identifying familial connections using samples containing DNA from two people and existing genealogical databases. Applying mathematical computation and machine learning algorithms, this deconvolution method is a breakthrough for forensic DNA casework, allowing analysts to decipher samples with DNA from two contributors. The DNA profiles relayed to a genealogical database can uncover familial matches. This tool for the emerging field of investigative genetic genealogy could help law enforcement identify missing persons and criminal suspects, especially in "cold" cases.

KEY FEATURES

- Identifies the sex (a helpful metric in identification of a person of interest) and DNA profile for each unknown contributor of two-person mixtures
- Enables searches for familial matches even to 3rd and 4th level relationships
- Demonstrated high performance in more than 500 simulated and 30 real experimental mixtures
- Provides a probability value associated with a predicted DNA profile

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Need

Since investigative genetic genealogy (IGG) was used to uncover the identity of the Golden State Killer in 2018, application of this forensic science has grown. An IGG search for familial matches is conducted only with a single-source DNA profile. However, many forensic casework samples contain DNA from multiple "contributors" who came in contact with a piece of evidence, leaving samples unresolved. Such a sample with mixed DNA must be deconvolved prior to its use for long-range familial searching, so a method to facilitate deconvolution enables generation of investigative leads from challenging samples.

Innovation

Lincoln Laboratory partnered with Verogen to develop a method to improve deconvolution. The method employs machine learning algorithms pioneered at Lincoln Laboratory to process the DNA data and produce results in the current standard Verogen ForenSeq Kintelligence sequencing format. During processing, multiple algorithms are applied to select two-person mixtures for evaluation; to identify the sex and DNA concentration of each contributor; and finally, to deconvolve each profile.

Demonstrations of the technique's performance, conducted using more than 500 simulated (representing 29 reported ethnicities) and 30 real experimental mixtures, successfully identifed the individual contributors in the two-person samples and enabled 3rd degree familial hits for most mixtures. The technique can be implemented with current IGG technologies, and its algorithms can be adapted for use by various devices.

Contributor	Concentration	Number of Mixtures	Sex (% Accuracy)	DNA Markers (% Deconvolved)	3rd Degree Familial Hit (% Accuracy)	4th Degree Familial Hit (% Accuracy)
Major	61%–98%	21	100%	95%	100%	56%
	55%–60%	2	100%	35%	72%	9%
Minor	5%–39%	18	83%	45%	61%	10%
	<5% or ≥40%	5	60%	34%	42%	5%

The table shows the mixture deconvolution results for sex prediction accuracy and percent of DNA markers deconvolved for the major and minor contributors across real experimental mixtures with various percent contributions ranging from 2 to 98% and ethnicity allowing for long-range familial hits (3rd and 4th degree relatives).

INTERESTED IN ACCESSING THIS TECHNOLOGY?

Contact the MIT Technology Licensing Office
https://tlo.mit.edu/
tlo-inquiries@mit.edu 617-253-6966

PATENT PENDING

INTERESTED IN WORKING WITH MIT LINCOLN LABORATORY?

https://www.ll.mit.edu/partner-us

Contact the Technology Ventures Office
tvo@ll.mit.edu

Approved for public release; distribution is unlimited. This material is based upon work supported by the CRADA under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the CRADA.