# Combining DNA Evidence for Greater Match Information

Mark W. Perlin, PhD, MD, PhD[a,*]

[a]*Cybergenetics, Pittsburgh, PA, USA*

**Abstract.** Statistical science studies multiple experiments in order to assess their implications and variation. Forensic science is often limited to a single experiment, and so cannot employ statistical inference. With DNA, however, there can be multiple items and assays from crime scene evidence. Since STR data have quantitative peak heights obtained from DNA sequencer signals, they can be statistically analyzed in a likelihood function. Moreover, multiple STR experiments can be mathematically combined using a joint likelihood function. Statistics tells us that using more data should yield more information. We find that computing with a joint likelihood function to combine DNA evidence can infer more identification information, as measured by the likelihood ratio match statistic.

*Keywords:* Forensic DNA, Joint Likelihood, Probabilistic Genotype, Likelihood Ratio, DNA Mixture, Peak Height

## 1. Introduction

Most fields of scientific enquiry routinely combine data from multiple experiments. These experiments can be repetitions drawn from one item, or involve different items entirely. The motivation is to elicit maximal information from an experimental design. The statistical mechanism is the joint likelihood function.

A likelihood function [1] mathematically quantifies how well alternative hypotheses explain a fixed data result. A joint likelihood function assesses these hypotheses on multiple data items simultaneously. Typically, the data are drawn from independent experiments. Therefore the joint likelihood simply multiplies together the likelihoods from separate experiments, jointly conditioned on a particular explanatory hypothesis [2].

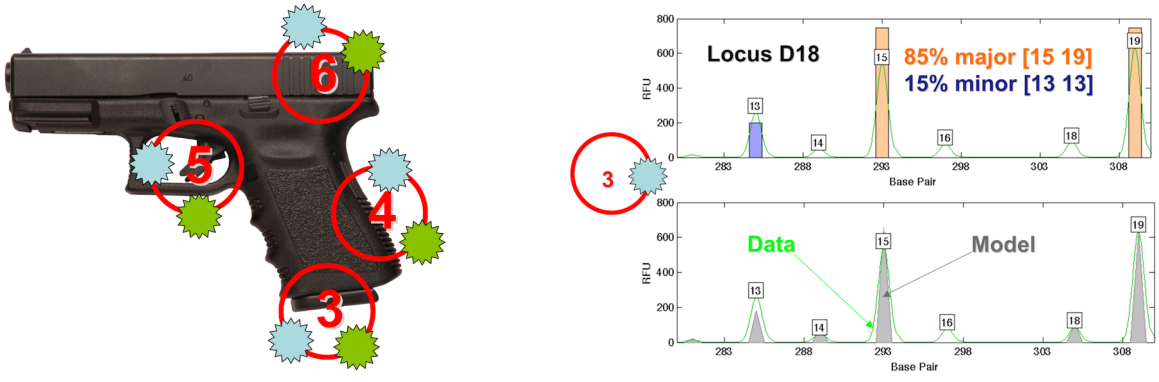
In forensic DNA science, human data interpretation is usually performed on data derived from only a single item. This practice is a consequence of thresholding quantitative peak height data into all-or-none qualitative allele possibilities, in order to simplify human review. Combining profiles after interpretation for "consensus" has little statistical foundation.

Quantitative computer interpretation, however, does not share these artificial limitations. It is therefore natural to mathematically preserve identification information by inferring a genotype using a joint likelihood function, examining all the independent data simultaneously [3, 4].

This paper describes the joint interpretation of DNA evidence. We show how likelihood functions can be used to rigorously explain DNA evidence, and how joint likelihood functions can combine evidence. We present data that shows how the number of assumed contributors affects the inferred result, and why appropriately constructed likelihood ratios cannot overstate the inferred DNA match information. We illustrate these concepts on representative DNA mixture cases and experiments [5, 6].

## 2. Materials and Methods

A handgun was swabbed in four locations, with each item amplified twice (Fig. 1A), yielding eight DNA data injections [5]. Human mixture review did not infer any result for the minor contributor.




**(A)** **(B)**

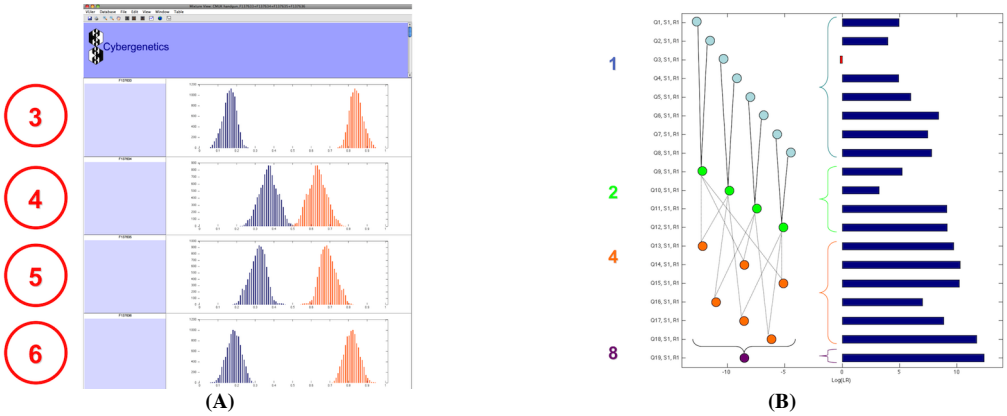
Fig 1. (A) A handgun swabbed in four locations (circles), each amplified twice (stars). (B) Likelihood function, showing mixture data and genotype model.

* Corresponding author. *Tel*: 001-412-683-3004. *Fax*: 001-412-683-3005.
*E-mail address*: perlin@cybgen.com. Cybergenetics, 160 North Craig Street, Suite 210, Pittsburgh, PA 15213, USA.

In a quantitative likelihood function, the computer must explain the observed peak heights (Fig. 1B). With uncertainty, the inferred genotype is a probability distribution over allele pairs. A joint likelihood function examines multiple data items, inferring genotypes that best explain all the observed evidence. Using more data reduces uncertainty, which sharpens the probability distribution and produces a higher likelihood ratio (LR).

## 3. Results

Using all four data items, the computer [4] inferred a unique genotype for the minor contributor, along with mixture weights (Fig. 2A). Jointly examining more data yielded more log(LR) identification information (Fig. 2B). While from just a single item the TrueAllele® computer could infer a useful LR of a hundred million, joint interpretation using all the data gave a LR of over a trillion (the full random match probability of a unique genotype).




**(A)** **(B)**

Fig 2. (A) DNA mixture weights for the four swabs. (B) DNA log(LR) match information, increasing as more data are combined together.

## 4. Conclusions

The purpose of forensic science is to preserve the full extent of identification information present in the evidence. Quantitative data permits the use of likelihood functions that can explain the data under alternative hypotheses. When there are multiple experiments, an entire data set can be jointly analyzed simultaneously by multiplying together these likelihood functions. This statistical approach provides a sound, well-founded procedure for combining evidence in order to preserve identification information. This paper applied the joint likelihood approach to DNA evidence having multiple items and amplifications. We established that using more DNA data can preserve more identification information.

## 5. Acknowledgements

William Allan, Meredith Clarke, Dr. Ria David and Matthew Legler at Cybergenetics assisted in this research.

## 6. Conflict of interest

Dr. Mark Perlin is a shareholder, officer and employee of Cybergenetics, an American company that develops the TrueAllele® Casework computer system that infers probabilistic genotypes from complex forensic DNA evidence (e.g., mixtures containing six contributors), and can combine DNA evidence using a joint likelihood function.

## 7. References

[1] Lindgren BW. Statistical Theory. Fourth ed. New York, NY: Chapman & Hall, 1993.
[2] Edwards AWF. Likelihood. Expanded ed. Baltimore: Johns Hopkins University, 1992.
[3] Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. *PLoS ONE*. 2009;4(12):e8327.
[4] Perlin MW, Legler MM, Spencer CE, et al. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 2011;56(November):in press.
[5] Perlin MW, Greenhalgh M. Scientific combination of DNA evidence: a handgun mixture in eight parts. *Twentieth International Symposium on the Forensic Sciences of the Australian and New Zealand Forensic Science Society*, September; Sydney, Australia. 2010.
[6] Ballantyne J, Perlin MW. DNA mixture deconvolution by binomial sampling of individual cells. *Eighth International Conference on Inference and Statistics (ICFIS)*; July 20; Seattle, WA, 2011.