Poster 291

Combining DNA Evidence for Greater Match Information

Dr. Mark W. Perlin, Cybergenetics, Pittsburgh, PA USA

ABSTRACT

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 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.

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.

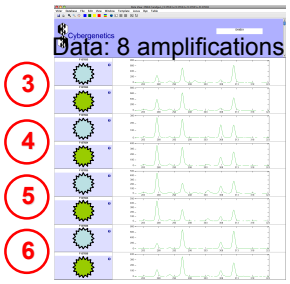
This poster describes the joint interpretation of DNA evidence. We show how likelihood functions can be used to rigorously explain DNA evidence, and how joint likelihood functions can combine evidence. We present data that shows how the number of assumed contributors affects the inferred result, and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

(Work done in collaboration with Matthew Greenhalgh of Orchid Cellmark in Abingdon, United Kingdom.)

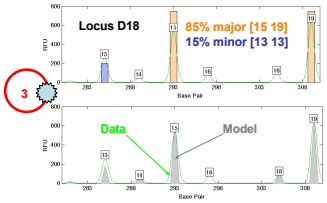
DNA Swabs

Duplicate Amplifications

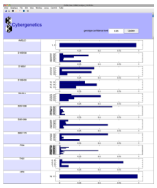
Data: 8 amplifications



Quantitative Likelihood Function



Computer Infers Minor Genotype

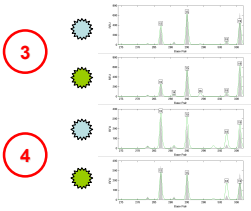


probability

Uncertain genotype
at every locus
< 100% probability
at multiple allele pairs

Finds match strength
log(LR) = 4.99
100,000-fold increase
0 → 4.99
over no result

Joint Likelihood Function



Locus D18

major [15 19]
minor [13 13]

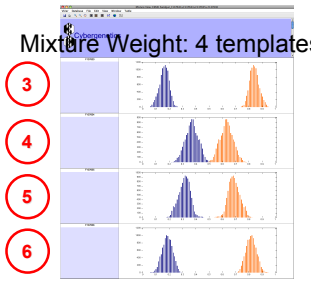
Infer Minor Genotype

probability

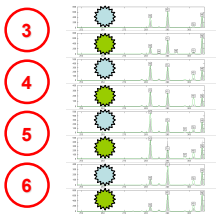
Higher genotype
certainty with
two amplifications of
two PCR templates

Greater match strength
log(LR) = 9.73
large increase
0 → 4.99 → 5.25 → 9.73

Mixture Weight: 4 templates

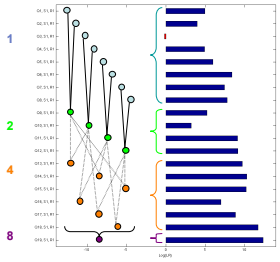


Joint Likelihood Function



Locus D18

major [15 19]
minor [13 13]



combinations information

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

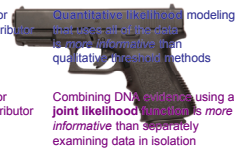
Row 2. In a **quantitative likelihood function**, the computer must explain the observed peak heights. With uncertainty, the inferred genotype is a probability distribution over allele pairs.

Row 3. 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).

Row 4. Using all four data items, the computer inferred a **unique genotype** for the minor contributor. Jointly examining more data yielded more log(LR) identification information. While from just a single item the TrueAllele® computer could infer a useful LR of a hundred million, its **joint interpretation** using all the data gave a LR over a trillion (the full random match probability).

Conclusions

Two mixture genotype examples



Major contributor: Quantitative likelihood modeling peak heights is more informative than qualitative threshold methods

Minor contributor: Combining DNA evidence using a **joint likelihood** function is *more informative* than separately examining data in isolation

Trillion-fold increase in identification information

Cybergenetics

Visit Cybergenetics website for papers and presentations: http://www.cybgen.com/information

Visit Cybergenetics ISFG booth to see a live computer demonstration of TrueAllele® Casework.

Cybergenetics © 2003-2011