Objective DNA Mixture Information in the Courtroom: Relevance, Reliability and Acceptance

Mark W. Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA

September 30, 2015

Presented in Arlington, Virginia on July 22, 2015 at
National Institute of Standards and Technology conference:

International Symposium on Forensic Science Error Management: Detection, Measurement and Mitigation

Corresponding Author:

Mark W. Perlin, PhD, MD, PhD
Chief Scientific Officer
Cybergenetics
160 North Craig Street, Suite 210
Pittsburgh, PA 15213 USA
(412) 683-3004
(412) 683-3005 FAX
perlin@cybgen.com

ABSTRACT

DNA mixtures arise when two or more people contribute their DNA to a biological sample. Data-simplifying thresholds fail to give accurate results when applied to complex mixture patterns. An entirely objective interpretation approach is to first separate out the genotypes of each mixture contributor, without ever seeing the subject, and only afterwards make a comparison.

Comparison of a separated evidence genotype with a subject's reference genotype, relative to a population, yields a match statistic. This likelihood ratio is a standard measure of information change based on observed evidence that addresses FRE 403 relevancy balancing. The reliability of objective genotype separation has been extensively tested. Such extensive testing, error rate determination, and scientific peer-review address FRE 702 and Daubert reliability factors.

Courts have accepted this extensively validated computer approach, with admissibility upheld at the appellate level. Separated genotypes provide results that juries find easy to understand. Objective DNA analysis elicits identification information from evidence, while rigorous validation establishes accuracy and error rates. Courts require solid science – extensively tested and empirically proven – to promote criminal justice, societal safety, and conviction integrity.

Table of Contents

Introduction

Deoxyribonucleic acid (DNA) mixtures arise when two or more people contribute their DNA to a biological sample. Mixtures are seen in sexual assault kits, homicide evidence, handguns and other “touch DNA” surfaces. With advances in detection technology, they have become the predominant form of DNA evidence in many crime laboratories. While DNA from one person is easy to interpret, mixture data has complex patterns comprising many allele peaks of varying height.

One person’s DNA produces either one allele peak, or two of similar height, so a height “threshold” is meaningful. But data-simplifying thresholds fail to give accurate results when applied to complex mixture patterns. Ten years ago, scientists at the National Institute of Standards and Technology (NIST) demonstrated a ten order-of-magnitude match statistic discrepancy between crime laboratories analyzing the same mixture data [1]. Mixture “inclusion” analysis tests whether a subject’s alleles are included in a set of (thresholded peak) alleles, but it is inherently subjective – the analyst sees the subject’s genotype during the analysis.

An entirely objective (and potentially more informative) approach is to first separate out the genotypes of each mixture contributor without ever seeing the subject, and only afterwards make a comparison. This can be accomplished by sophisticated computing that considers many thousands of genotype alternatives, and how well their additive combinations explain the quantitative data [2]. Multiple possibilities for a contributor genotype are assigned probabilities. Faithful modeling of the laboratory process can yield genotypes that accurately preserve DNA identification information.

Comparison of a separated evidence genotype with a subject's reference genotype, relative to a population, gives a match statistic. This statistic is a simple ratio – the probability of genotype match divided by the random match probability. The statistic is also a likelihood ratio (LR), or Bayes factor (BF), which is a standard measure of information change based on observed evidence.

Mathematically, the LR is probative because it assesses how evidence data affects a hypothesis (i.e., whether the subject contributed their DNA to the mixture). The LR's assessment is also non-prejudicial, because (as a BF) the ratio factors out prior belief about the hypothesis. Thus genotype separation addresses Federal Rules of Evidence (FRE) 403 relevancy balancing.

The reliability of objective genotype separation has been extensively tested for at least one such system. Dozens of independent and developmental validation studies have been conducted, with seven peer-reviewed TrueAllele® publications. These studies use the LR as an objective information measure to assess the method's sensitivity (true positives), specificity (false positives) and reproducibility (close numbers). This extensive testing, error rate determination, and scientific peer-review address FRE 702 and Daubert reliability factors.

Courts have accepted this extensively validated computer method, which has withstood Daubert and Frye challenges in six states. Admissibility has been upheld at the appellate level. Separated genotypes provide results that are easy to understand.

Objective DNA analysis elicits identification information from evidence. Validation establishes accuracy and error rates. Courts require solid science – extensively tested and empirically proven – to promote criminal justice, societal safety, and conviction

integrity. This paper describes DNA mixtures, and how to objectively interpret them, focusing on relevance, reliability, and acceptance.

Case Example

We examine DNA mixture evidence in a Baltimore trial of Nelson Clifford; the author was an expert witness for the prosecution. Arguing consent, Clifford had been acquitted of sexual offenses on four previous occasions [3]. In this fifth case, mixtures were found on articles of clothing – a green shirt and a belt. The forensic question was: “Did suspect Nelson Clifford contribute his DNA to the victim’s clothing?”

A mixture sample contains DNA from two or more people. Figure 1 shows a relatively large amount of DNA from one person (blue) who has a 6,8 allele pair, a second person (orange) who is homozygous for allele 7, and a third person (green) with a 7,9 allele pair. The additive combination of these relative DNA amounts produces a data signature for this particular biological mixture.

Bayes law

Bayes law lets us reach meaningful conclusions from a small amount of data. Bayes uses this data to update belief. The probability law is 250 years old [4], but has gained considerable traction in the last 50 years with the advent of digital computing [5].

Bayes begins with a prior probability (brown, right side) of what we believe before we see data (Figure 2). We examine data through a likelihood function that describes

how well a hypothesis explains the data, giving a probability number (green, middle). All hypotheses are considered, determining how the data updates our belief (blue, left). The result is a posterior probability, our final belief after we have observed the data.

Genotype modeling is the application of Bayes law to genetic identification (Figure 3). We begin with a random genotype (brown, right) of probabilities for about 100 different allele pair possibilities at each locus. The quantitative data is then examined, usually for short tandem repeat (STR) data [6].

A computer considers all genotype possibilities, along with variables such as stutter, degraded DNA, variances, and other parameters. After examining the data, we derive a new genotype probability. This result represents our belief in the different genotype values for each contributor at every genetic locus.

STR data

Bayesian analysis starts with the data. We have STR genetic data comprised of quantitative peak heights, shown for the green shirt mixture at locus TH01 (Figure 4). There is a pattern of taller peaks at alleles 6 and 8, and lower peaks at 7 and 9.

It is important to use all of the data. Specifically:

(a) The *amounts* of the DNA matter, expressed as peak heights that reflect the relative quantities of each allele in the biological sample.

(b) The *pattern* of high and low peaks matter, as these patterns can be explained by different genotype hypotheses of allele pair quantities and their artifacts.

(c) The peak *variation* is needed for modeling variance parameters; there can be

dozens of these parameters in a DNA mixture problem. For example, the 6 and 8 peaks here represent roughly the same amounts of DNA contributed by one person, but we see variation in their (unequal) peak heights.

Genotype separation

A likelihood function helps separate out the genotypes of each contributor to a mixture. The likelihood explains the genotyping data mathematically. Shown is one such explanation, out of many thousands that were considered (Figure 5). There is a major amount of a first 6,8 allele pair (blue), a minor amount of a second homozygote allele pair at allele 7 (orange), and a minor amount of DNA for third allele pair 7,9 (green).

Adding up these three different allele pairs forms a pattern, where the heights of those cumulative allele quantities are (to a first approximation) near the peak heights of the observed data. Since this pattern explains the data well, it has a relatively high likelihood and thus confers higher probability to each of the contributor genotypes.

A separated contributor genotype is shown in Figure 6. The locus vs. contributor table (center) lists 13 genetic loci, with TH01 in the first row, followed by another 12 loci. Each of the three assumed contributors has a separate column. There are thus 39 locus contributors (13 loci × 3 contributors), each with its own separated genotype.

The bar graph (blue) shows one such genotype, here for a minor contributor at the TH01 locus. Out of a hundred or so possible TH01 allele pairs, the STR data has focused probability onto about a half dozen of these possibilities (x-axis). The probability scale is also shown (y-axis). Each bar gives the posterior probability of seeing an allele

pair (for this minor contributor at TH01), after having seen the STR mixture data.

This objective genotyping procedure is unbiased by the suspect's genotype; the computer is not given that reference information, only the mixture data. Moreover, the process is unbiased by a human analyst subjectively selecting data peaks. Data is entered into a machine, and then analyzed automatically. This mechanization facilitates workflow and productivity, but also ensures objectivity.

This process infers a separated genotype for each contributor at every locus. These objectively derived mixture genotypes are recorded on a computer's hard drive. We can now use these separated genotypes to calculate a DNA match statistic, relative to the suspect.

Relevance and match

Our forensic comparison goal is to assess the strength of match. We consider FRE 403, which governs the relevance of evidence. We want to assess the identification hypothesis "Did the suspect contribute his DNA to the mixture?" The legal role of relevance is to balance the probative force of DNA evidence against the danger of unfair prejudice to the defendant (Figure 7).

The likelihood ratio conducts this balancing mathematically. The LR is a form of Bayes theorem for a single hypothesis [7]. It quantifies the question "To what extent does the evidence increase or decrease strength in the identification hypothesis?"

The LR has a numerator (blue) that measures the extent to which the hypothesis is impacted upon by data. This numerator is inherently *probative*, since it centers on

how evidence affects the hypothesis. The denominator (brown) states the initial *prejudicial* odds of the identification hypothesis before seeing data. In dividing numerator by denominator, the LR factors out the prior prejudice from the evidentiary probative force.

After applying Bayes theorem and some algebra, we can calculate the likelihood ratio through genotype posterior probability [8]. At the defendant's genotype, we simply divide the probability *after* having seen data by the probability *before* seeing data. That is how genotypes give us match statistics. They provide a way of using DNA data to calculate a likelihood ratio for the identification hypothesis.

Match simplification

Separated genotypes are much easier to understand than unmixed STR data. With a separated genotype, mixture comparison is like random match probability (RMP), the standard DNA statistic involving just one genotype. We ask the question, "To what extent does the evidence match the suspect more (or less) than a random person?"

The graph in Figure 8 shows the same *posterior* genotype probability distribution (blue bars) as before – the separated contributor at the TH01 locus after the data has been seen. Now also shown (brown bars) are a half dozen (out of a hundred) allele pair possibilities having *prior* genotype probabilities for a random person in the population – the prior gives the chance that we are seeing a match purely by coincidence.

With these posterior (blue) and prior (brown) genotype probabilities, we can make a statistical comparison with anyone's genotype. In this case, the genotype of the

defendant happens to be a 7,9. We therefore focus our attention on that allele pair (red rectangle), looking at the ratio of posterior (blue bar) to prior (brown bar) probability at 7,9. This ratio of 47% to 13% equals 3.62, the value of the likelihood ratio at TH01.

The LR is the posterior genotype probability at the suspect's genotype, divided by the probability of a coincidence. We see that the numerator's 47% is less than the full 100%. A 100% numerator over a 13% denominator would be the simple RMP match statistic. But a DNA mixture introduces match uncertainty, so we must consider that reduced strength of match in the numerator, in addition to the usual genotype rarity in the denominator. Using separated genotypes, the LR is just the old RMP but with a reduced numerator; this idea is easy to understand and explain in court.

The match statistic is shown for each locus by a horizontal bar (Figure 9). The 13 loci are listed from top to bottom. Since STR genetic loci are independent, we can multiply these values together to calculate the joint LR. Stated in plain language, a match between the shirt and Nelson Clifford is 182,000 times more probable than coincidence.

Exclusionary power

Also of interest is the exclusionary power of a matching genotype. Comparing the contributor genotype (over all loci) with 10,000 random genotypes, we obtain a bell shaped curve of match statistics (Figure 10). This non-contributor distribution describes the match information (on a logarithmic scale) for someone who did *not* contribute their DNA to the mixture. The logarithmic mean is around –10, for an average exclusionary

power of 1 over 10 billion; for a non-contributor, a coincidence is far more probable than an evidence-based match. The standard deviation (yellow bar) is around three log units.

From this non-contributor distribution (Figure 11), we can calculate an error rate for the match statistic (purple math). The LR is 182 thousand, which has a $\log_{10}$(LR) of 5.25. The normal distribution's z-score for this log(LR) value is 5.02, or five standard deviations to the right (yellow bar). That deviation has a p-value tail probability of 2.53×10^{-7}. Therefore, the chance of observing a non-contributing individual with a LR of at least 182 thousand (i.e., a false inclusion) is 1 in 4 million.

Case outcome

Figure 12 shows a separated DNA mixture. TrueAllele separated the green shirt mixture into three genotypes: 11%, 82% and 7% contributors. These genotypes were objectively inferred, without examination bias from the suspect or some other reference. Following genotype separation, comparisons were made to three references (victim, elimination and Clifford), yielding match statistics to each of the three mixture contributors.

In this fifth Clifford case, the jury convicted him of third degree sex offense [9]. "Only DNA connected Clifford to the masked man who terrorized" his victims [10]. The defendant's prior sex offense was considered when he was sentenced to over 30 years in prison.

TrueAllele Validation

TrueAllele has been extensively validated in dozens of studies conducted by Cybergenetics and crime labs. Four peer-reviewed studies were performed on laboratory-synthesized data of known composition – mixtures that are made in the laboratory [11-14]. Three other peer-reviewed studies were done on casework samples, which have more realistic data complexity [15-17]. Both types of studies should be done when thoroughly validating a DNA mixture interpretation method.

A recent TrueAllele validation paper appeared in the *Journal of Forensic Sciences*. The study was conducted with co-author Kevin Miller in collaboration with the Kern County crime laboratory in Bakersfield, California. Entitled "TrueAllele genotype identification on DNA mixtures containing up to five unknown contributors," the study employed a realistic randomized mixture design.

The Kern paper reported seven main results. The "contributor sufficiency" axis examined how changing the computer's assumed number of contributors affects the match statistic. This axis showed that once there are a sufficient number of assumed unknown contributors, TrueAllele's match statistic does not materially change.

For example, suppose there are actually three contributors in a DNA mixture. When the computer conducts separate runs assuming three, four, or five unknown contributors, the statistical match results will be essentially the same. Therefore, TrueAllele does not need to know the true number of contributors. Three other axes of interest were specificity, sensitivity, and reproducibility.

Specificity

Specificity validation studies are helpful in court. Figure 13 shows the distribution of log(LR) values for comparisons made between separated mixture genotypes and random genotypes. Millions of genotype comparisons were made, and the log(LR) values were recorded. The mixtures contained 2, 3, 4 or 5 unknown contributors.

The LR data are shown on a logarithmic scale. Zero log(LR) means there is no information (blue vertical line). As the number of contributors increases (from 2, to 3, 4 or 5), specificity (or exclusionary power) decreases. With five contributors in low-template DNA, the average is over one in a billion.

Specificity data can be used to develop a table of false positive events, as was done in this validation study. The table provides false inclusion error rate information. When asked in court, “What is the chance of seeing a false inclusion when the match statistic is a thousand?” the response can be an accurate numerical estimate. With log(LR) non-contributor data collected and tabulated, the error rate becomes a definite probability, whether one in a thousand or one in a trillion.

Sensitivity

Sensitivity examines to what extent a method can detect someone who actually contributed DNA to a mixture. As we increase mixture complexity from two to five contributors, the contributor distribution shifts leftwards towards less identification information (Figure 14). However, even with five contributors, and very low DNA

quantities, TrueAllele successfully made most of the identifications.

Reproducibility

We assessed TrueAllele reproducibility by running the program twice on the same data under the same conditions. Each point on the scatterplot in Figure 15 shows log(LR) values from two independent computer runs on one mixture. The points line up nicely along a 45° angle, showing that the replicated numbers are essentially the same. Reproducibility was measured using a within-group standard deviation statistic, and found to be well under a log unit, regardless of DNA quantity or contributor number.

Reliability

Reliability is important for the admissibility of scientific evidence. Expert evidence should be based on reliable methods that have been reliably applied to sufficient data. Daubert admissibility factors include whether a method is testable, has an associated error rate, has undergone peer-review, and is generally accepted in the relevant scientific community. The Frye standard considers only general acceptance.

TrueAllele has been admitted after Daubert challenge in Louisiana and Ohio. The system has withstood Frye challenges in California, New York, Pennsylvania and Virginia. Internationally, TrueAllele has successfully weathered “voir dire” challenges in Australia and the United Kingdom.

Acceptance

TrueAllele acceptance is widespread. Judicial acceptance has been facilitated by validation studies. The first TrueAllele case was tried six years ago in Pennsylvania, which led to an appellate precedent in that state [18].

TrueAllele has since been used in hundreds of criminal cases, and in over half of the states in the United States. TrueAllele experts appear mainly for the prosecution, but also testify for the defense. Five crime labs now regularly use TrueAllele in their criminal casework, with California having started in 2013 [19].

The main impact of TrueAllele is in bringing DNA evidence back into criminal cases. Past and current crime laboratory interpretation guidelines discard most mixtures as “inconclusive,” or assign weak statistics. This information loss precludes the evidence from being heard in court. TrueAllele restores mixtures as viable DNA evidence, with guilty pleas a common outcome.

Conclusions

Objective genotyping can help eliminate examination bias. When a calculator doesn’t know the comparison profiles, interpretation can’t be directed toward a desired answer. After separating out genotypes from a mixture, they can be compared against any number of people (one, two, ten, or a entire database).

Identification information (the likelihood ratio logarithm) is a standard information statistic. The log(LR) quantifies DNA information in a case, as well as in a validation

study. The LR condenses the many aspects of genotype comparison into a single number. Scientific LR validation can help establish accuracy, applicability, and error rates. These assessments aid in understanding DNA mixture evidence, and how to use or explain it in court.

There are untested mixture interpretation methods. For example, the manual combined probability of inclusion (CPI) method has enjoyed widespread use for 15 years [20]. CPI is a probabilistic genotyping approach based on a very simple likelihood function, one that does not make much use of the data [21]. CPI accuracy has not been assessed, even though its reliability has been questioned [22,23]. Validation is needed to demonstrate CPI's relevance and reliability.

Courts need solid forensic science that has been empirically proven. Untested DNA mixture statistics should not be offered as reliable evidence. With objective and reliable science, better data interpretation achieves better criminal justice, helping to protect society and maintain conviction integrity.

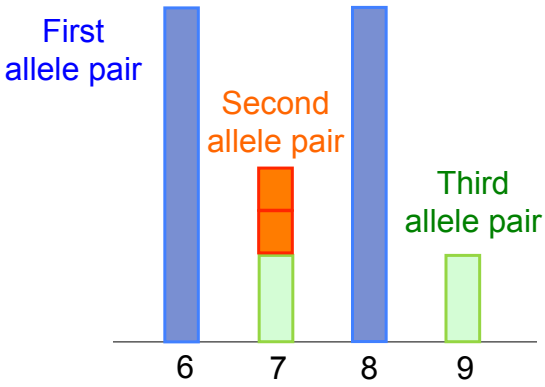
References

1. Butler JM, Kline MC (2005) NIST Mixture Interpretation Interlaboratory Study 2005 (MIX05), Poster #56. Promega's Sixteenth International Symposium on Human Identification. Grapevine, TX.
2. Perlin MW, Szabady B (2001) Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. J Forensic Sci 46: 1372-1377.
3. Anderson J, Duncan I (2013) Man charged in multiple sexual assault cases acquitted a fourth time. The Baltimore Sun.
4. Bayes T, Price R (1763) An essay towards solving a problem in the doctrine of chances. Phil Trans 53: 370-418.
5. O'Hagan A, Forster J (2004) Bayesian Inference. New York: John Wiley & Sons.
6. Butler JM (2005) Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. New York: Academic Press.
7. Good IJ (1950) Probability and the Weighing of Evidence. London: Griffin.
8. Perlin MW (2010) Explaining the likelihood ratio in DNA mixture interpretation. Promega's Twenty First International Symposium on Human Identification. San Antonio, TX.
9. George J, Fenton J (2015) Jury convicts sex offender in fifth trial. The Baltimore Sun.
10. Fenton J (2015) Repeat sex offender apologizes, gets 30 years in prison. The Baltimore Sun.
11. Perlin MW, Sinelnikov A (2009) An information gap in DNA evidence interpretation. PLoS ONE 4: e8327.
12. Ballantyne J, Hanson EK, Perlin MW (2013) DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. Sci Justice 53: 103-114.
13. Perlin MW, Hornyak J, Dickover R, Sugimoto G, Miller K (2014) Assessing TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors (poster). AAFS 66th Annual Scientific Meeting. Seattle, WA: American Academy of Forensic Sciences.
14. Greenspoon SA, Schiermeier-Wood L, Jenkins BA (2015) Establishing the limits of TrueAllele® Casework: a validation study. J Forensic Sci 60: 1263-1276.
15. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, et al. (2011) Validating TrueAllele® DNA mixture interpretation. J Forensic Sci 56: 1430-1447.
16. Perlin MW, Belrose JL, Duceman BW (2013) New York State TrueAllele® Casework validation study. J Forensic Sci 58: 1458-1466.
17. Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S (2014) TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. PLoS ONE 9: e92837.
18. (2011) Commonwealth of Pennsylvania v. Kevin James Foley. Superior Court of Pennsylvania.
19. Perlin MW, Miller KW (2014) Kern County resolves the DNA mixture crisis. Forensic Magazine 11: 8-12.

20. Scientific Working Group on DNA Analysis Methods (SWGDAM) (2000) Short Tandem Repeat (STR) interpretation guidelines. Forensic Science Communications 2.
21. Perlin MW (2010) Inclusion probability is a likelihood ratio: implications for DNA mixtures (poster #85). Promega's Twenty First International Symposium on Human Identification. San Antonio, TX.
22. Curran JM, Buckleton J (2010) Inclusion probabilities and dropout. J Forensic Sci 55: 1171–1173.
23. Brenner CH (2011) The mythical “exclusion” method for analyzing DNA mixtures – does it make any sense at all? (A111). AAFS 63rd Annual Scientific Meeting. Chicago, IL: American Academy of Forensic Sciences. pp. 79.

1. DNA mixture

Two or more people contribute their DNA to a sample



2. Bayes law

Use data to update belief (1762)

Prob(hypothesis | data) proportional to
Prob(data | hypothesis) x Prob(hypothesis)

New belief, **after** seeing data	How well hypothesis **explains** data	Old belief, **before** seeing data
posterior	*likelihood*	*prior*

3. Genotype modeling

Apply Bayes law to genetic identification

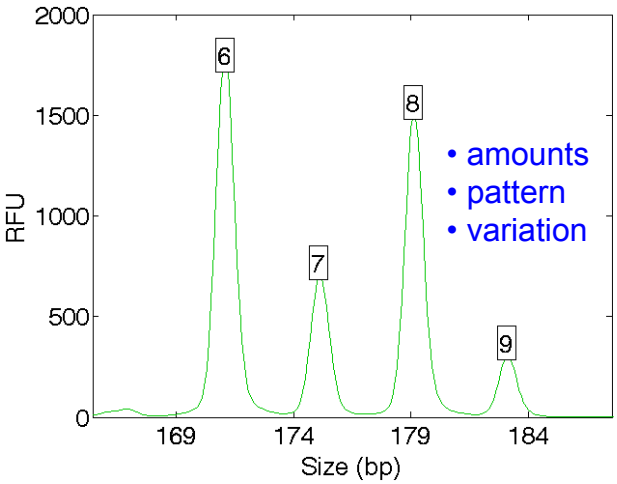
Prob(genotype | data) proportional to
Prob(data | genotype) x Prob(genotype)

New genotype probability, **after** seeing data	How well genotype choice **explains** data	Old genotype probability, **before** seeing data
posterior	*likelihood*	*prior*

Probabilistic genotyping

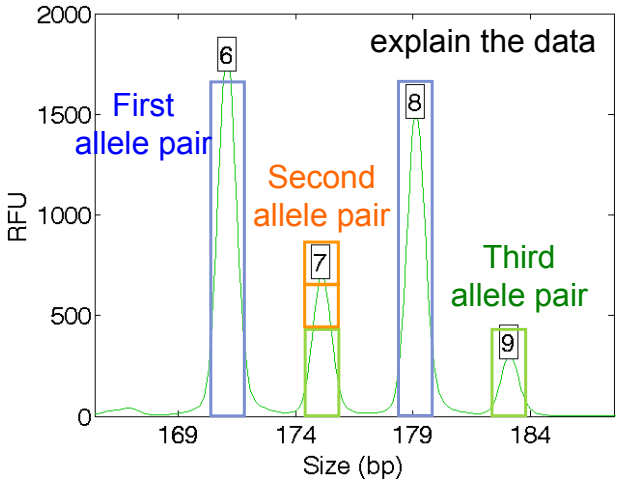
4. Genetic data

Quantitative peak heights at locus TH01



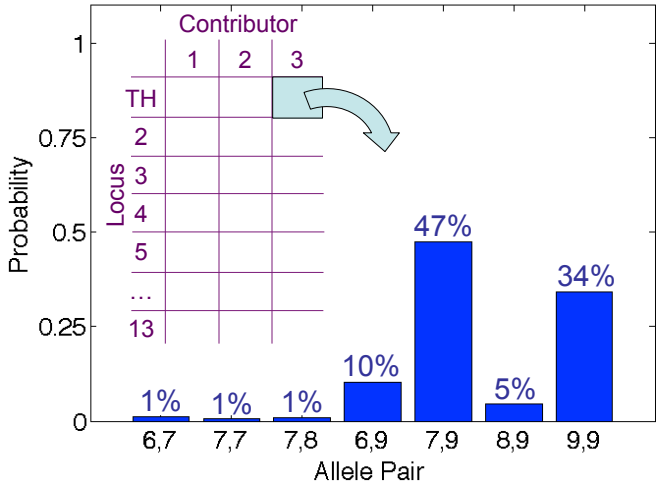
5. Separate genotypes

Consider *every possible* genotype (Bayes)



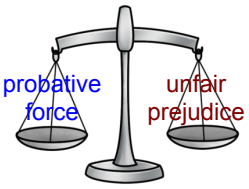
6. Separated genotype

Objective, unbiased – doesn't know suspect's genotype



7. Relevance (FRE 403)

Hypothesis = "suspect contributed his DNA"



likelihood ratio (LR)
is Bayes law
for a hypothesis

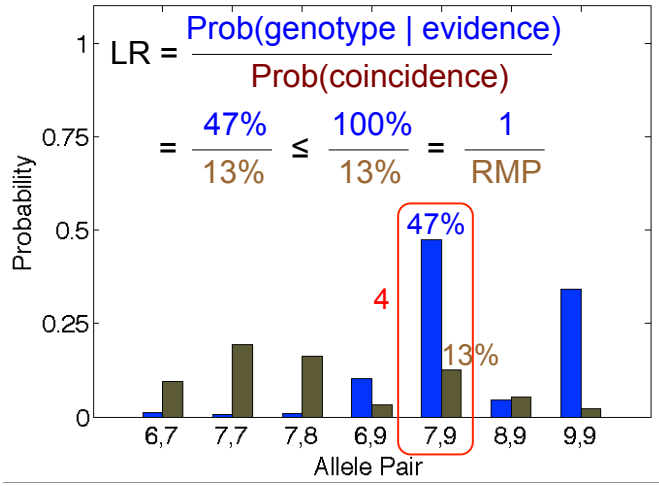
Probative

$$LR = \frac{\text{Odds(hypothesis | data)}}{\text{Odds(hypothesis)}} = \frac{\text{Prob(genotype | data)}}{\text{Prob(genotype)}}$$

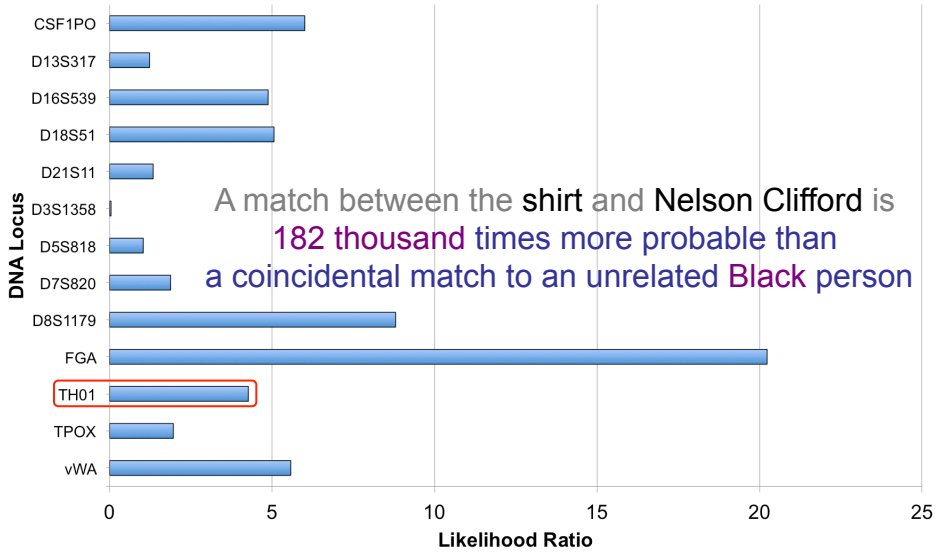
Non-prejudicial

8. Match statistic is simple

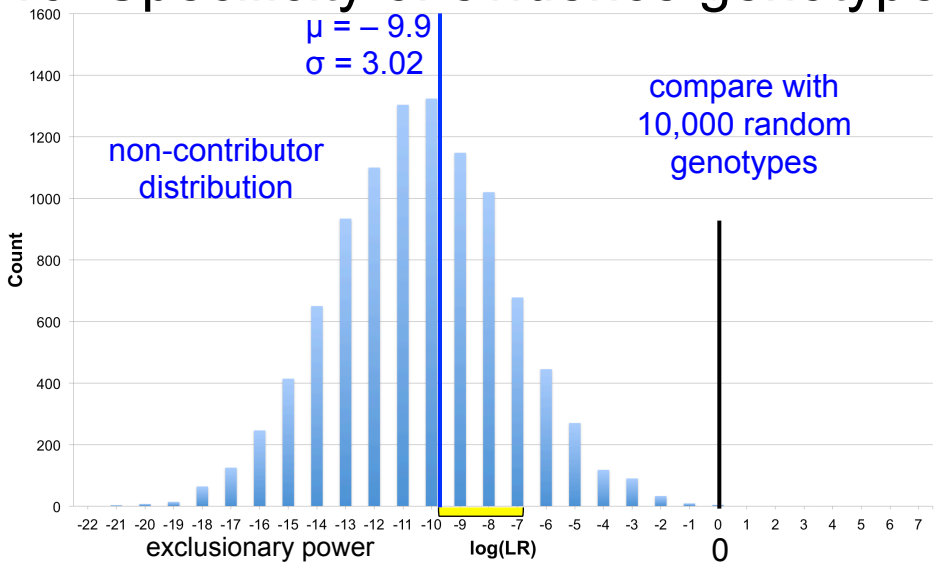
Suspect matches evidence more than random person



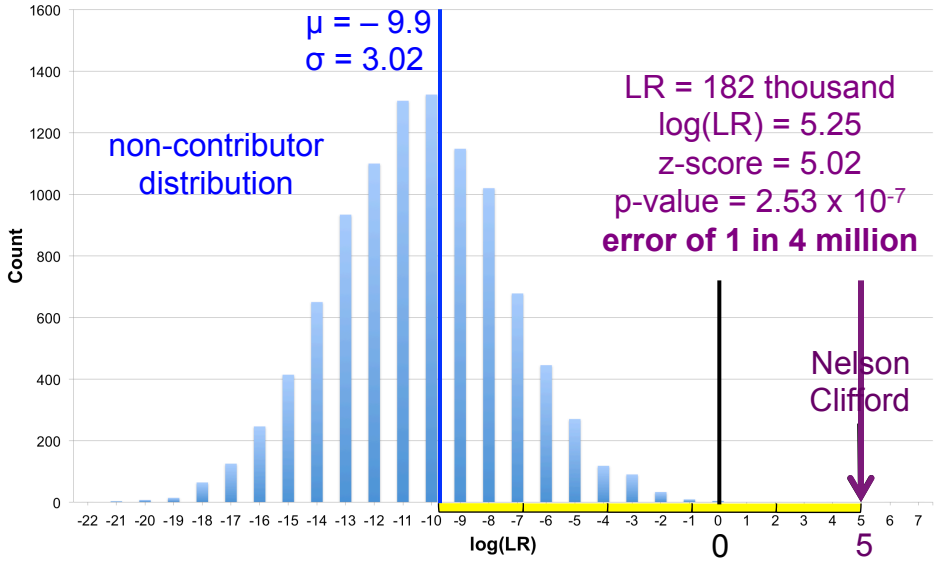
9. Match statistic at all loci



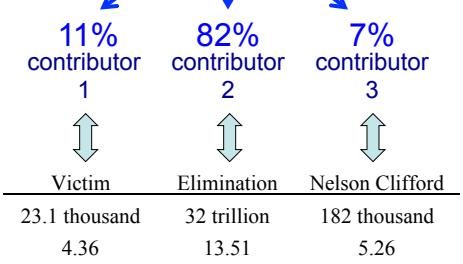
10. Specificity of evidence genotype



11. Error rate for match statistic



12. Separated DNA mixture



	Victim	Elimination	Nelson Clifford
LR	23.1 thousand	32 trillion	182 thousand
log(LR)	4.36	13.51	5.26

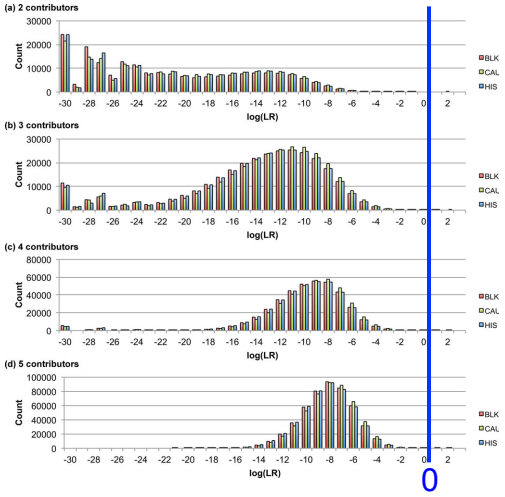
13. Specificity

2

3

4

5



- low-template DNA
- compare millions

- exclusionary power
- contributor number

- false positive table
- error rate in court

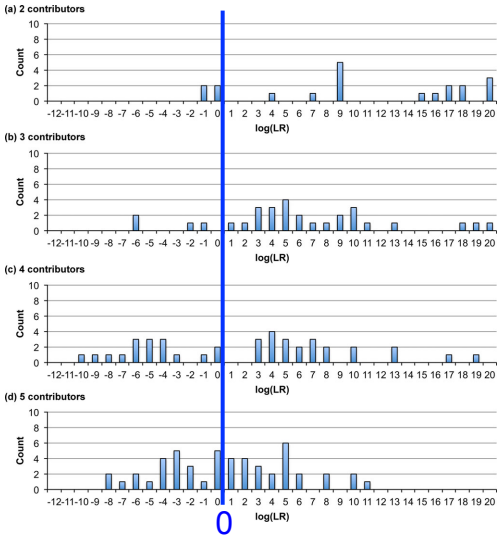
14. Sensitivity

2

3

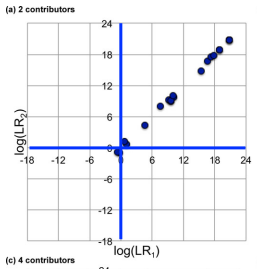
4

5

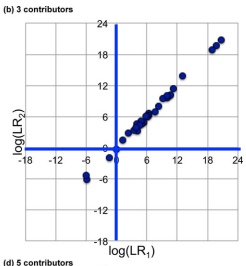


15. Reproducibility

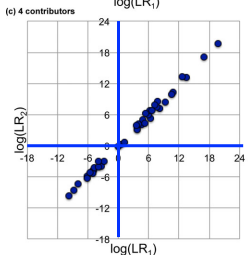
2



3



4



5