Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 2 (2008) 91–103
FSI GENETICS
www.elsevier.com/locate/fsig

Interpretation of complex DNA profiles using empirical models and a method to measure their robustness

Peter Gill [a,*], James Curran [b], Cedric Neumann [a], Amanda Kirkham [a], Tim Clayton [c], Jonathan Whitaker [c], Jim Lambert [c]

[a] Forensic Science Service, Trident Court, 2960 Solihull Parkway, Solihull B37 7YN, UK
[b] Department of Statistics, The University of Auckland, Private Bag 92019, Auckland, New Zealand
[c] Forensic Science Service, Sandbeck Way, Audby Lane, West Yorkshire LS22 7DN, UK

Received 3 May 2007; received in revised form 3 September 2007; accepted 9 October 2007

Abstract

A new methodology is presented in order to report complex DNA profiles. We have brought together a number of different theories in order to devise a new protocol to interpret complex cases using likelihood ratios. The calculations are designed to be highly conservative and are widely applicable. We apply a low copy number (LCN) interpretation framework, which includes the probabilities of dropout and contamination, to 'conventional' DNA cases. In conventional casework, stutters often compromise calculations when they are observed with the same height as a minor contributor to a mixture. Stutters cannot be distinguished from minor alleles. We compensate by treating them as real alleles and including them in the calculation. By increasing the number of potential contributors to the DNA profile, we can account for the extra alleles that result. We propose that the likelihood ratio is qualified with additional robustness parameters to indicate the probability of misleading evidence in favour of the prosecution, under the assumption that a random man was a contributor instead of the suspect. To do this we apply a new kind of case-specific 'Tippett' test. Although the method is complex, we suggest a 'user-friendly' way to explain the results to a court. The method is easily extended to carry out ranked likelihood ratio (LR) searches for suspects in national DNA databases.
© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Mixtures; Low copy number; Tippett test; Expert system; *LoComatioN*; Likelihood ratio

1. Introduction

1.1. Elimination of the 'inconclusive' DNA profile

Using traditional methods, it is only possible to report a DNA profile that actually matches the suspect in whole or in part. Consequently, the probative value given always has a likelihood ratio (LR) greater than one. However, DNA profiles are often ambiguous—they may be partial, with alleles missing under prosecution (H_p) propositions; they may be mixtures; stutters may interfere with the interpretation. Every DNA scientist will routinely make decisions on whether to report a profile in the context of missing alleles, or additional alleles in the profile that do not match the suspect. Expert opinion is used to carry out the assessment, but this can lead to reporting inconsistencies where some scientists may apply a probative value to a result, whereas others may be more 'cautious' and provide an inconclusive result that neither includes nor excludes the suspect.

For complex profiles, this means that a traditional calculation can only be carried out provided that simplifying assumptions are made—for example, if the profile is partial then we must assume that dropout has occurred under the prosecution hypothesis. But, if the suspect genotype is *ab* and the crime stain profile is *a*, then the numerator probability is less than one, and a traditional LR calculation may not be conservative [1,2], especially if the peak in the crime stain is sufficiently large such that the probability of dropout is effectively zero ($\Pr(D) \approx 0$) [2]. If the DNA profile is complex, or if there are several bands that do not match the suspect reference profile, then extra contributors or contamination (probabilistically measured by $\Pr(C)$ [1]) must also be considered. The judgement 'call' may be to report the DNA profile as 'inconclusive', using a phrase such as "no meaningful

* Corresponding author.
E-mail address: dnapgill@compuserve.com (P. Gill).

1872-4973/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2007.10.160

statistic can be applied''. This is just is another way of saying that the profile is too complicated to interpret.

The question follows whether the LR result is robust. To investigate robustness we utilise 'Tippett' tests. These tests originate from Tippett et al. [3] and were originally applied to paint. Evett and Weir [4], pp. 213–215 applied the concept to early examples of DNA analysis and a detailed summary of the philosophy of the Tippett test is provided by Buckleton et al. [5], pp. 188–191. Previously, Tippett tests have been used as a general test of robustness of a particular technique or method applied across randomly generated cases (typically 1000 separate tests).

We have adopted previous ideas and developed them into an alternative strategy to test the robustness of the LR of a *specific* case. This is the first application to complex (partial) mixture STR analysis. In our strategy, judgement calls are still required by the reporting scientist, particularly in the area of formulating a probability of dropout, but we have formalised the process succinctly. A framework is provided to incorporate all alleles into the calculation without the need to specifically assign all of them to specific contributors or to stutter artefacts—this is especially important if a minor DNA profile is of evidential significance since the probative alleles and alleles in stutter positions may not be reliably distinguished. Also we provide a new method to determine whether proposed improvements to models are effective or worthwhile. The result is an entirely new concept that applies insight into the robustness of the likelihood ratio measurement itself.

Consequently, there is no longer a need for an 'inconclusive' category for reporting purposes that is based on perceived complexity of the result. In principle any profile (with any number of potential contributors) can be probabilistically evaluated against any set of hypotheses.

Because we have removed much of the subjective nature of the interpretation of the DNA profile itself, this firmly shifts the focus of the courtroom debate into a consideration of the relevant alternative pairs of hypotheses that form the likelihood ratio. Hypotheses are formed by consideration of the casework circumstances, and by a consideration of the DNA profiles themselves. Often this process is not straightforward and it may be appropriate to consider multiple pairs of propositions. Gill et al. [6] show how simplification of this process can be achieved by reporting a minimum likelihood ratio of a defined set of proposition pairs. Hitherto, the complexity of carrying out such calculations has precluded this kind of approach to considering evidence.

We show that there is no need for an 'inconclusive' category for reporting purposes that is based on perceived complexity of the result because in principle any profile can be probabilistically evaluated against any set of hypotheses.

1.2. How robust is the answer?

Once a likelihood ratio has been calculated, then case-specific Tippett tests provide insight into the robustness of the test itself. Specifically, we are most interested in answering the question: ''*if the suspect is not a contributor to the crime-stain, how likely is it that a LR of similar magnitude would be reported if a random man was the contributor*''? We use computer simulation to address this issue. Instead of calculating the LR relative to the suspect and conditioned profiles under the prosecution hypothesis (H_p), we replaced these profiles with random men – which are the proposition under the defence hypothesis (H_d). We demonstrated that the resulting LRs were substantially less than one and the chance that random man would give probative evidence was negligible in the examples we discuss.

It is also of interest to consider the alternative scenario ''*if the suspect truly was the contributor to the crime stain, how often will the LR be less than one*''? How often will the evidential strength favour the defence hypothesis when the suspect really is the perpetrator? This is also tested using simulation. Conditioning on all of the contributors under H_p a new evidence profile is generated by conditioning on $\Pr(D)$ and $\Pr(C)$. The LR is calculated and the simulation repeated an arbitrarily large number of times (1000 in this paper). We demonstrated that the LRs generated were above one and the actual case-specific calculated LRs were within their respective simulated ranges.

We show that these principles can be usefully applied to complex cases that hitherto could not have been reported beforehand.

2. The role of expert judgement in relation to DNA profiling evidence

Empirical guidelines are routinely used in classical DNA mixture interpretation to designate alleles and to make decisions on the numbers of contributors. Expert decisions tend to be binary, i.e. a probability of an event is either zero or one. For example:

Stutter determination: in reality, a minor peak that is in a stutter position (−4 bp from a major allele) is either a stutter or an allele. Alternatively, it may be a mixture of both stutter and allele. The expert will usually make a definitive decision that is often based on the size of the adjacent (parent) allele. A common guideline that is used is the 15% threshold, which has been derived from experimental observations that the stutters tend to be less than 15% the size of the parent allele. Taking an unambiguous, unmixed sample as an example: if a peak in stutter position is less than 15% of the size of the adjacent (+4 bp) allele, then it is designated as a stutter, i.e. $\Pr(\mathrm{St}) \approx 1$ [7], and it does not feature further in the calculations. Similarly, if the peak exceeds 15% of the size of the parent allele then the peak will be designated as an allele, i.e. $\Pr(\mathrm{St}) \approx 0$. Other information may be used to qualify this decision.

Heterozygote pairing: Heterozygote balance measures the relative sizes, in terms of peak height (Ht) or peak area, of two alleles. Hence heterozygote balance can be defined as

$$\mathrm{Hb} = \frac{\mathrm{Ht}_{\mathrm{smallest}}}{\mathrm{Ht}_{\mathrm{largest}}} \times 100\%$$

Download English Version:

https://daneshyari.com/en/article/99256

Download Persian Version:

https://daneshyari.com/article/99256

Daneshyari.com