Forensic Science International: Genetics 16 (2015) 13–16

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Do low template DNA profiles have useful quantitative data?

Duncan Taylor [a,b,*], John Buckleton [c]

[a] Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia
[b] School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia
[c] ESR, Private Bag 92021, Auckland 1142, New Zealand

ARTICLE INFO

Article history:
Received 5 September 2014
Received in revised form 10 October 2014
Accepted 5 November 2014

Keywords:
DNA profile interpretation
Mixtures
Likelihood ratios
Low template
Continuous models

ABSTRACT

A set of low template mixed DNA profiles with known ground truths was examined using software that utilised peak heights (STRmix^{TM} V2.3) and an adapted version that did not use peak heights and mimicked models based on a drop-out probability [1,2] (known as semi-continuous or 'drop' models) (STRmix^{TM} lite). The use of peak heights increased the LR when H_p was true in the vast majority of cases. The effect was most notable at moderate template levels but was also present at quite low template levels. There is no level at which we can say that height information is totally uninformative. Even at the lowest levels the bulk of the data show some improvement from the inclusion of peak height information.

1. Introduction

Methods for evaluating DNA profiles have benefitted from recent improvements in modelling and software [2–8]. This has allowed the interpretation of many more profiles and the generation of a corresponding relevant match statistic, namely a likelihood ratio (LR) [3,7,9]. Previously Taylor [10] demonstrated that the LR generated by a continuous model [3] trended towards 1 as the template was reduced or as the number of contributors increased. This was true whether a true or false donor to the mixture was considered.

The question has been legitimately asked whether there is a point where the quantitative data present in peak height information becomes no more informative than merely the presence or absence of the peak.

There has been a view that profiles can be low template and that it is these low template profiles where peak height is of limited or no value. However for mixed DNA profiles it is likely that the contributors will be present in different template amounts. Hence each of, say, four contributors could be more or less low in template level. In addition nearly all casework profiles have a downward slope with respect to molecular weight. This is often termed a degradation slope. What this means is that it may be simplistic to refer to profiles as high or low template. Many profiles will exhibit a range of template estimates dependent on molecular weight. This observation has been made previously [11]. What this means is that the method used to interpret such profiles, which are prevalent in casework, must be able to interpret information that is very likely to range in template from high to low within the same profile or contributing component to a mixture.

There are three lemmas that can be considered useful at this point:

(1) Adding correct and relevant information to a calculation can only increase the ability to distinguish between a true and false proposition
(2) If the amount of information provided to a calculation is decreased, at some point the ability to distinguish between true and false propositions is entirely lost
(3) If a contributing profile can be determined unambiguously then additional information (such as peak height data) will not improve the ability of a calculation to distinguish between a true and false proposition

The hypothesis we wish to consider is: At low template the stochastic effects are such that the addition of peak height information to a calculation provides negligible additional ability to distinguish between true and false propositions.

Inline with lemma 3 there is little point trialling DNA profiles whose contributing genotypes can be determined unambiguously, as we know the addition of peak height information in these

---

[*] Corresponding author at: Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia. Tel.: +61 8 8226 7700; fax: +61 8 8226 7777.
E-mail address: Duncan.Taylor@sa.gov.au (D. Taylor).

http://dx.doi.org/10.1016/j.fsigen.2014.11.001
1872-4973/

instances will have no effect. What is needed are a set of low template mixed DNA profiles with ground truth known that have been analysed with and without the use of peak height information.

To this end we trial complex mixed DNA profiles with a range of input DNA, where the mixture proportions and donor profiles were known, to assess at what point peak height information no longer benefits an LR calculation, and specifically the ability of the LR to distinguish between known contributors (H_p true) and known non-contributors (H_d true).

2. Method

A range of four person mixtures produced in GlobalFiler (Thermo Fisher Scientific), as per manufacturer's instructions. Amplification fragments were resolved using the ABI PRISM® 3130xl Genetic Analyser and analysed in Genemapper® ID-X to obtain peak height information for each profile. These mixtures are samples 22–31 from [10], amplified in triplicate (note that there are only two replicates of sample 23 rather than three). We reproduce the relevant mixture information from [10] in Table 1.

As in [10] profiles were analysed down to 30 rfu.

STRmix^{TM} version 2.3 (http://strmix.esr.cri.nz) was reconfigured to ignore peak height information. A probability of dropout, Pr(D), was required and was applied to each instance of a peak dropout. $1 - \Pr(D)$ was applied to each instance of non-dropout. The value of Pr(D) was determined by including it as a parameter in the model, which sampled from its posterior with a flat prior by the Markov chain. This is effectively maximum likelihood estimation for Pr(D) for every profile during its analysis. Note that a single Pr(D) was applied across all contributors to the profile. This reconfigured STRmix^{TM} product was termed STRmix^{TM} lite and clones the performance of semi-continuous ('drop') models in a manner that is as close as possible to the normal functioning of STRmix. This was done, rather than using separate software, so that all factors other than the one of interest (peak height) would remain constant between the two experiments.

In both STRmix^{TM} V2.3 and STRmix^{TM} lite a uniform probability for allelic drop-in of 0.0017 was used inline with laboratory observations. To calculate LRs each combination of three individuals was assumed for each four person mixed DNA profile, meaning from the 29 profiles, 116 analyses were carried out and compared to a POI using the propositions:

H_p: The POI, contributor A, contributor B and contributor C are the sources of DNA
H_d: Contributor A, contributor B and contributor C and an unknown individual are the sources of DNA

where POI, A, B and C were combinations of contributors $\mathbf{C_{1\text{-}4}}$.

LRs were calculated using an in-house self-declared Caucasian GlobalFiler database and using the product rule.

Table 1
Mixture proportions and PCR setup.

Tubes	Mixture proportions for contributor				Total DNA added to PCR (pg)
	C_1	C_2	C_3	C_4	
22	0.25	0.25	0.25	0.25	400
23	0.25	0.25	0.25	0.25	200
24	0.25	0.25	0.25	0.25	50
25	0.25	0.25	0.25	0.25	20
26	0.25	0.25	0.25	0.25	10
27	0.40	0.30	0.20	0.10	400
28	0.40	0.30	0.20	0.10	200
29	0.40	0.30	0.20	0.10	50
30	0.40	0.30	0.20	0.10	20
31	0.40	0.30	0.20	0.10	10

3. Results

In Fig. 1 we give the improvement in the $\log_{10}$(LR) when peak height data is included in the analysis and plot against the input DNA of individual contributors. As input template level is not directly available from an electropherogram we also give the improvement against average allelic peak height of each profile.

Peaks below about 300 rfu are indicative of low template. Even profiles where the average peak height is 300 rfu often have low template components.

In Fig. 2 we plot the log of LR produced by STRmix^{TM} lite vs the log of LR produced by STRmix^{TM} V2.3. This allows an investigation of the benefit of peak heights for profiles that would produce a LR below 10^9 if peak height was not used.

In Table 2 we provide the results of H_d true testing (for an explanation of the H_d true test concept see Evett et al. [12] or Gill and Haned [13]) for sample 24, assuming C_2, C_3 and C_4 in both STRmix^{TM} (left) and STRmix^{TM} lite (right). H_d true tests are when non-donors are compared to a profile in order to generate an LR. In the propositions given in Section 2 we replace the POI with a DNA profile that has been randomly simulated in accordance with expectations from population allele frequencies. A large number of H_d true tests can be performed in order to give a series of 'diagnostics' about the profile analysis and in particular the performance of the models used to analyse the profile data. In Table 2 we provide the following values:

Simulations: The number of randomly simulated profiles that were compared to the evidence DNA profile
H_p true LR: The value of the LR obtained when compared with the known contributor

And for H_d true comparisons

p ('1 in'): *p* is the proportion of H_d true tests that yielded an LR at least as big as the LR obtained from the known contributor. Values give are the inverse of *p* so that they can be directly compared to the size of the H_p true LR
LR = 0: The percentage of simulations that resulted in an LR of zero being obtained
LR ≥ 1: The percentage of simulations that resulted in an LR that favoured the inclusion of the randomly simulated non-donor. These have classically considered as 'false inclusions'.
Average LR: The average value of all the H_d true LRs. Theory predicts that this value should be one as explained in Good [14] (quoting Turing).

4. Discussion

The results presented in Fig. 2 demonstrate a strong advantage in using peak height information down to very low levels. We see that the $\log_{10}$(LR)s with and without peak height information converge towards 0 (LR = 1) as the information in the profile diminishes. Hence both approaches are correctly reporting that the profile becomes uninformative at extremely low template.

Most (110/116) instances of including peak height information yielded a higher LR when compared to known contributors. The greatest effect can be seen for results at high LRs. These differences are of less practical importance.

However for LRs less than one billion in STRmix^{TM} lite (i.e. to the left of the vertical line in Fig. 2) there is still significant increase in LRs when peak height information is included. In many instances this is four or five orders of magnitude. There is no level at which we can say that height information is totally uninformative. Even