Forensic Science International: Genetics Supplement Series 5 (2015) e37–e38

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

A simple unconstrained semi-continuous model for calculating likelihood ratios for complex DNA mixtures

Daniel Russell[a,*], William Christensen[b], Todd Lindsey[b]

[a] Oklahoma City Police Department (OCPD) Forensic DNA Laboratory, Oklahoma City, Oklahoma, United States
[b] Oklahoma State Bureau of Investigation (OSBI) Forensic Science Center, Edmond, Oklahoma, United States

ARTICLE INFO

Article history:
Received 20 August 2015
Accepted 7 September 2015
Available online 9 September 2015

Keywords:
Unconstrained likelihood ratio (UCLR)
Likelihood ratio
DNA
Statistics
Mixtures

ABSTRACT

Methods for calculating likelihood ratios to provide weight for complex DNA mixtures have been developed, but factors including the complexity of the equations and assumptions about the possible number of contributors to a profile have prevented some laboratories from embracing these methods. By producing and studying complex DNA mixture profiles and approaching interpretation using a minimal number of assumptions, we have developed a semi-continuous method for calculating likelihood ratios. The unconstrained likelihood ratio (UCLR) model for providing statistical weight to inclusions is simple and can be performed without making assumptions about the number of contributors. Additionally, we are able to establish and test the limits of the model, providing greater assurance that false inclusions will not occur by chance.

Published by Elsevier Ireland Ltd.

1. Introduction

Advancements in DNA technology have contributed to an increase in the number of samples which involve complex DNA mixtures. Multiple mathematical models and programs have been created to perform calculations that attempt to show the strength of the evidence as it relates to the DNA profiles of the individuals that are potential contributors to these complex mixtures [1,2]. Most models require assumptions to be made by the analyst or the computer program in order to assess the profile. A common assumption regarding the number of contributors to these complex mixture profiles can be problematic, as the methods used to assess the number of contributors in complex mixture profiles have been shown to be unreliable [3,4]. We provide data further supporting the hypothesis that the number of contributors cannot be accurately determined and use the data to assess a semi-continuous method for calculating a likelihood ratio to provide weight to an inclusion, which reduces the number of assumptions an analyst must make.

2. Materials and methods

2.1. Mixture study preparation

A series of DNA mixtures were prepared using known contributors in varying ratios to evaluate the ability to determine the number of contributors present in a mixture sample, as well as to evaluate the capabilities of the proposed unconstrained likelihood ratio (UCLR) model. Ten varying ratios of 3, 4, and 5 person mixtures were prepared and amplified with the Applied Biosystems® Identifier Plus® Amplification kit and analyzed with the Applied Biosystems® 3500 Genetic Analyzer.

In addition, ten mixtures for each number of contributors (3–5) were created by random selection from a database of 39 DNA profiles, and these “virtual” mixtures (with no drop-out) were compared to known non-contributors from the same database to help determine limitations for complex mixture interpretation.

2.2. Establishing the unconstrained likelihood ratio (UCLR) model

The unconstrained likelihood ratio (UCLR) model calculates a single, static likelihood ratio that provides an estimation of the strength of the probability of a given individual's DNA profile being present in a DNA mixture versus the probability of anyone else's profile being represented in the mixture profile. The UCLR model works with minimal assumptions being made by the analyst, and the goal of the research is to make the model simple and amenable

* Corresponding author at: OCPD Laboratory, 616 Colcord, Oklahoma City, OK, 73102.
E-mail address: daniel.russell@okc.gov (D. Russell).

http://dx.doi.org/10.1016/j.fsigss.2015.09.015
1875-1768/Published by Elsevier Ireland Ltd.

$$\left[\left(\sum_{i=1}^{n} p_i\right)^2 (D_0)\right] + \left[\left(\left(\sum_{i=1}^{n} p_i\right)^2 + \left(2\sum_{i=1}^{n} p_i - 2\left(\sum_{i=1}^{n} p_i\right)^2\right)\right)(D_1)\right] + \left[\left(\left(\sum_{i=1}^{n} p_i\right)^2 + \left(2\sum_{i=1}^{n} p_i - 2\left(\sum_{i=1}^{n} p_i\right)^2\right) + \left(1 - \left(\sum_{i=1}^{n} p_i\right)\right)^2\right)(D_2)\right] =$$

Equation 1 $$\left[\left(\sum_{i=1}^{n} p_i\right)^2 (D_0)\right] + \left[\left(2\sum_{i=1}^{n} p_i - \left(\sum_{i=1}^{n} p_i\right)^2\right)(D_1)\right] + D_2$$

Fig. 1. The overall equation for the denominator.

to validation. The model is semi-continuous, and peak heights are only used for drop-out calculations. Drop-in is not accounted for due to its rare occurrence in laboratories with well-established thresholds. The denominator of the UCLR model is the probability of finding a random individual with the required drop-out for them to be represented in the observed profile. An abbreviated simplification of the overall equation is shown in Fig. 1.

The numerator of the UCLR model is the probability that a given individual's profile could be represented in the observed DNA profile from the questioned samples given that the individual is a contributor. The overall probability of the numerator depends on the probability of drop-out. At each locus, the numerator would be $(D_0 + D_1 + D_2) = 1$ for fully represented profiles, $D_1 + D_2$ for partially represented profiles, and D_2 for un-represented profiles.

2.3. Model behavior

From the equation in Fig. 1 it can be observed that if $D_0 = 1$, the inverse of the CPI results for a fully represented individual which is the maximum likelihood ratio allowed for any mixture. If no alleles are observed at a locus, then $D_2 = 1$, and all of the other terms are equal to zero. The resulting likelihood ratio would equal 1, so the locus will offer no information. As the sum of allelic frequencies and the probability of drop-out at a locus increases, the overall LR decreases.

3. Results and discussion

3.1. Results of mixture analysis

The allele counting method, which uses the maximum alleles at any one locus, was used to calculate the number of contributors for each of the prepared mixtures. Eighty percent of the three person mixtures were accurately assessed as a three person mixture, but twenty percent were inaccurately assessed as having only two contributors. All of the four and five person mixtures were inaccurately assessed as having one fewer contributor than the true composition of the mixture. The same method was applied to the randomly generated virtual mixtures. The three person virtual mixtures were accurately assessed, but seventy percent of the four person virtual mixtures were inaccurately assessed as only having three contributors. All of the five person virtual mixtures were inaccurately assessed by the allele counting method with eighty percent assessed as four contributors and twenty percent assessed as only three contributors.

Comparisons were made between the non-contributor DNA profiles from the database to the prepared mixture sets as well as to the randomly generated virtual mixtures. As the number of contributors increases, random inclusions can be made when allowing as little as 1 or 2 alleles to be absent from the DNA mixture. In the five person virtual mixtures, it was observed that one non-contributor profile was fully represented in one DNA mixture profile.

3.2. Determining limits of the UCLR model and results of testing

Any statistical model will have limits and it is important to identify those limits when the developing a model. Establishing limits allows for only relevant data to be reported and facilitates assessment through validation. Parameters that have been identified in this study that can be used to help determine the UCLR model's limits include the overall denominator, the total of observed allele frequencies, normalization factors to bring calculated results in line with empirical results, and the ratio of the maximum LR possible to the calculated LR.

Data produced from 150 mixtures of 3, 4, and 5 contributors were each analyzed with the UCLR model using 60,000 randomly generated profiles. These profiles were compared to the mixtures producing a total of 9,000,000 comparisons of varying complexity. Reasonable weights were associated with the comparisons that were made, and the parameters listed above provided additional relevance to the LR's that were calculated. The UCLR was capable of making relevant inclusions for the correct contributors to the mixtures, but further testing will be necessary to fully assess the model.

4. Concluding remarks

The number of contributors in a complex mixture often cannot be accurately identified. As the complexity of the mixture increases, very little drop-out can be tolerated before it is possible for random inclusions to occur by chance. The UCLR model provides a simple calculation that provides appropriate weight to inclusions while also limiting the chances of false inclusions without the need to make multiple assumptions that cannot be supported by the observed data.

Conflict of interest

None.

Acknowledgements

We thank the staff at the Oklahoma State Bureau of Investigation for providing resources to allow for the completion of this project. All of the data used in this work was generated at the OSBI Forensic Science Center using their supplies and equipment. We would also like to thank Dr. John Butler for his insight and guidance on this project.

References

[1] M.W. Perlin, J.M. Hornyak, G. Sugimoto, K.W.P. Miller, TrueAllele® genotype identification on dna mixtures containing up to five unknown contributors, J. Forensic Sci. 60 (4) (2015) 857–868.
[2] P. Gill, J. Curran, C. Neumann, A. Kirkham, T. Clayton, J. Whitaker, J. Lambert, Interpretation of complex DNA profiles using empirical models and a method to measure their robustness, Forensic Sci. Int. Genet. 2 (2008) 91–103.
[3] J.-A. Bright, J. Curran, J. Buckleton, The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation, Forensic Sci. Int. Genet. 12 (2014) 208–214.
[4] H. Haned, L. Pene, J. Lobry, A. Dufour, D. Pontier, Estimating the number of contributors to forensic dna mixtures: does maximum likelihood perform better than maximum allele count? J. Forensic Sci. 56 (1) (2011) 23–28.