

# Development of new peak-height models for a continuous method of mixture interpretation



Sho Manabe<sup>a,b,\*</sup>, Yuya Hamano<sup>a,c</sup>, Chihiro Kawai<sup>a</sup>, Chie Morimoto<sup>a</sup>, Keiji Tamaki<sup>a</sup>

<sup>a</sup> Department of Forensic Medicine, Kyoto University Graduate School of Medicine, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

<sup>b</sup> Research Fellow of Japan Society for the Promotion of Science, Japan

<sup>c</sup> Forensic Science Laboratory, Kyoto Prefectural Police Headquarters, 85-3, 85-4, Yabunouchi-cho, Kamigyo-ku, Kyoto 602-8550, Japan

## ARTICLE INFO

### Article history:

Received 26 August 2015

Accepted 12 September 2015

Available online 25 September 2015

### Keywords:

Forensic DNA typing

Mixture interpretation

Continuous model

Locus-specific amplification efficiency

Heterozygote balance

Stutter ratio

## ABSTRACT

DNA mixture interpretation based on a continuous model is an effective strategy for calculating rigorous likelihood ratios using peak heights and considering stochastic effects. Such a model would require the elucidation of various biological parameters affecting the expected peak heights. In the present study, we estimated the distributions of locus-specific amplification efficiency, heterozygote balance, and stutter ratio in 15 commercially available short tandem repeat (STR) loci using 234 single-source DNA samples. Our data suggested that the locus-specific amplification efficiency followed a normal distribution, whereas the heterozygote balance followed a log-normal distribution for each locus. We modeled log-normal distributions for stutter ratios with allele-specific mean values, which exhibited a positive correlation with allele repeat numbers. However, with the D8S1179, D21S11, and D2S1338 loci, the log-normal distribution did not fit our data because of the complex repeat structures involved. Therefore, an alternative model for each of these three loci will need to be incorporated into a software program based on a continuous model.

© 2015 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

DNA mixture interpretation using short tandem repeat (STR) loci is based on a binary model that does not account for peak-height information in DNA profiles. In recent years, some countries have begun to use continuous models that use the peak heights, including stochastic effects (e.g., allele drop-out), to calculate rigorous likelihood ratios [1]. This model can avoid some of the criticisms regarding the subjectivity of DNA mixture interpretation.

Appropriate use of the continuous model requires application of some biological parameters affecting the probability of the peak heights given all the possible genotype combinations of the contributors. In the present study, we estimated the distributions of three parameters (i.e., locus-specific amplification efficiency,

heterozygote balance, and stutter ratio) in 15 commercially available STR loci using single-source DNA samples.

## 2. Materials and methods

### 2.1. STR typing

Buccal samples were collected from 276 individuals using a Buccal DNA Collector (Bode Technology, Lorton, VA). Extraction from buccal cells was performed using BioRobot<sup>®</sup> EZ1 (Qiagen, Hilden, Germany) found in the EZ1 DNA investigator kit according to standard protocols. Extracted DNA was amplified using an AmpF/STR<sup>®</sup> Identifiler<sup>®</sup> Plus PCR Amplification Kit (Life Technologies, Carlsbad, CA) following the manufacturer's instructions. PCR products were then analyzed on an ABI 3130xl Genetic Analyzer (Life Technologies) and data were analyzed using GeneMapper<sup>™</sup> ID version 3.2.1 (Life Technologies) using 30 relative fluorescence units (RFU) as the limit of detection. We excluded 42 DNA samples from our estimation of the distributions of the parameters because of primer-binding site mutations ( $n=4$ ), tri-allelic patterns ( $n=1$ ), off-ladder alleles ( $n=5$ ), and pull-up peaks stacked on the stutter

\* Corresponding author at: Department of Forensic Medicine, Kyoto University Graduate School of Medicine, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. Fax: +81 75 761 9591.

E-mail address: [manabe@fp.med.kyoto-u.ac.jp](mailto:manabe@fp.med.kyoto-u.ac.jp) (S. Manabe).

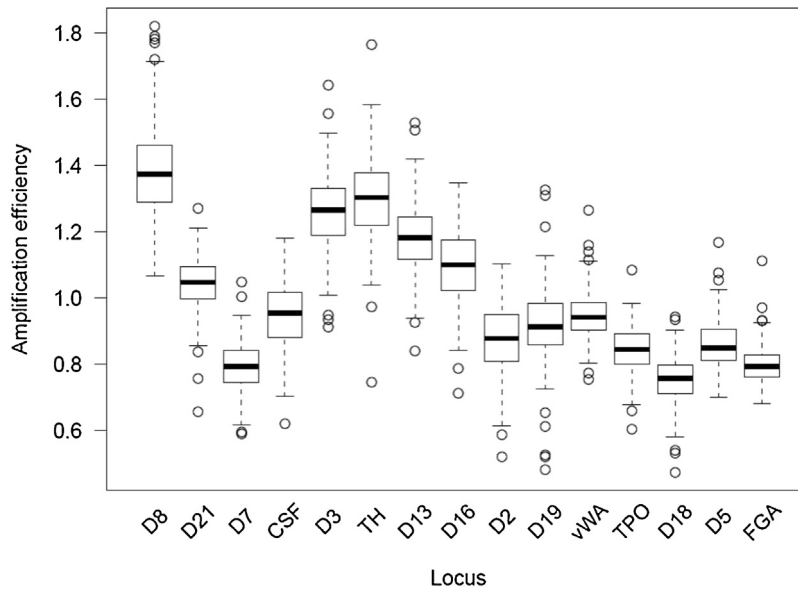


Fig. 1. Distributions of the  $A_l$  values at each locus.

peaks ( $n = 33$ , including one sample in which we also detected an off-ladder allele). Finally, we used 234 DNA samples to estimate the distributions of the three parameters.

2.2. Calculation of the three parameters

Locus specific amplification efficiency ( $A_l$ ) was defined as follows:

$$A_l = \frac{T_l}{\bar{T}}$$

where  $T_l$  denotes the sum of all allelic and stutter peak heights in locus  $l$  ( $l = 1, 2, \dots, L$ ), and  $\bar{T}$  denotes the mean value of  $T_l$  (i.e.,

$$\bar{T} = \frac{\sum_{l=1}^L T_l}{L}$$

We calculated  $A_l$  values in each locus of all 234 experimental profiles.

Heterozygote balance (Hb) was defined as follows:

$$Hb = \frac{O_{a'-1} + O_{a'}}{O_{a-1} + O_a}$$

where  $O_a$  refers to the height of the low-molecular-weight allele, and  $O_{a'}$  is the height of the high molecular weight allele.  $O_{a-1}$  and  $O_{a'-1}$  denote stutter peak heights of alleles  $a$  and  $a'$ , respectively. To implement the Hb distribution in a software program based on a continuous model, the effects of stutter ratio should be eliminated from the Hb calculation. Thus, we defined Hb as the ratio of total allelic products (i.e., sum of the allele peak height and stutter peak height), not the ratio of allele peak heights. If the stutter peak of the high-molecular-weight allele was masked by the low-molecular-weight allele (i.e.,  $a = a' - 1$ ), we did not calculate the Hb value.

Stutter ratio (SR) was calculated as follows:

$$SR = \frac{O_{a-1}}{O_a}$$

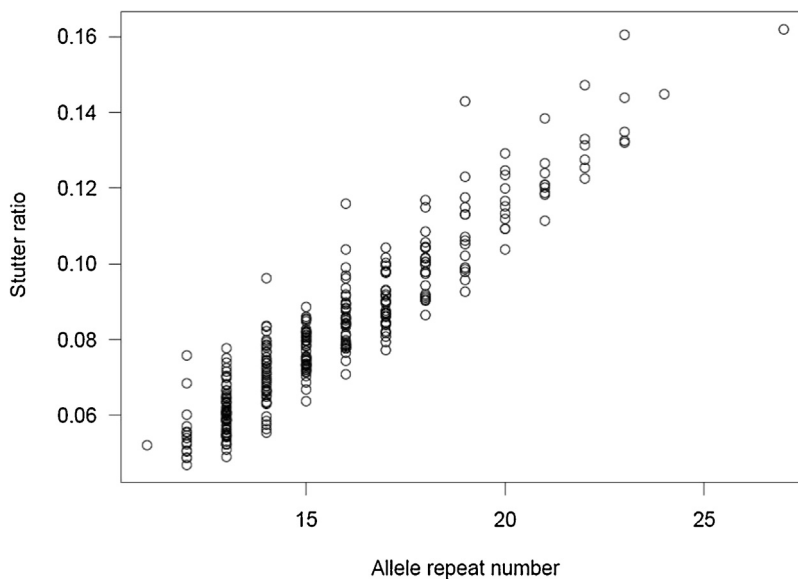


Fig. 2. Plots of SR vs. allele repeat number in D18S51.

where  $O_a$  refers to the height of the allele  $a$ , and  $O_{a-1}$  refers to the stutter peak height of allele  $a$ . If the stutter position was the same as another allelic position in a heterozygous locus, we did not calculate the SR value.

### 3. Results and discussion

Fig. 1. shows the distribution of the  $A_l$  values in each locus. The D8S1179 locus had the highest median value of  $A_l$  (1.37), whereas the D18S51 locus had the lowest median value of  $A_l$  (0.757). We assumed that  $A_l$  followed a normal distribution because the data were symmetrically distributed. The assumption was checked using quantile–quantile (Q–Q) plots. The Q–Q plots showed good agreement with the observed  $A_l$  values.

In the same way, we investigated the distribution of the Hb values at each locus. The median values were nearly equal to one for all loci. We assumed that Hb followed a log-normal distribution because the data were symmetrically distributed in the logarithmic scale. The Q–Q plots of the log-normal distribution showed good agreement with the observed Hb values.

Fig. 2 shows the distributions of SR values in D18S51 locus. The SR values were positively correlated with allele repeat numbers. We observed this trend in 11 loci but not in for D8S1179, D21S11, TH01, and D2S1338. As previously reported, we assumed that the SR values in the 11 loci followed a log-normal distribution with allele-specific mean values [2]. The assumption resulted in good prediction of the observed SR values using the Q–Q plots.

In the TH01 locus, the SR values of allele 9.3 were close to those of allele 6. Bright et al. showed that the longest uninterrupted stretch (LUS) is a more reliable predictor of SR than the allele repeat number [2]. The LUS value of allele 9.3 is 6 according to a

previous sequence analysis [3]. The TH01 locus also followed the log-normal distribution as the allele repeat number of allele 9.3 was 6.

However, LUS values could not be determined for D8S1179, D21S11, and D2S1338. For example, there are two types of repeat structures in the D8S1179 locus (i.e.,  $[TCTA]_a$  and TCTA TCTG  $[TCTA]_{a-2}$  for allele  $a$ ) [3]. Therefore, for these three loci, an alternative model is required and must be incorporated into a software program based on a continuous model.

#### Role of funding

None.

#### Conflict of interest

None.

#### Acknowledgment

This work was supported by a Grant-in-Aid for JSPS Fellows (JSPS KAKENHI grant number 14J03372).

#### References

- [1] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, *Forensic Sci. Int. Genet.* 7 (2013) 516–528.
- [2] J.-A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, *Forensic Sci. Int. Genet.* 7 (2013) 296–304.
- [3] J.M., Butler, D.J., Reeder, Short tandem repeat DNA internet database. Available from: [www.cstl.nist.gov/biotech/strbase](http://www.cstl.nist.gov/biotech/strbase).