Legal Medicine 19 (2016) 77–82

Contents lists available at ScienceDirect

Legal Medicine

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

Estimating allele dropout probabilities by logistic regression: Assessments using Applied Biosystems 3500xL and 3130*xl* Genetic Analyzers with various commercially available human identification kits

Shota Inokuchi [a,b,c,*], Tetsushi Kitayama [b], Koji Fujii [b], Hiroaki Nakahara [b], Hiroaki Nakanishi [a], Kazuyuki Saito [a], Natsuko Mizuno [b], Kazumasa Sekiguchi [b]

[a] Department of Forensic Medicine, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo, Tokyo 113-8421, Japan
[b] National Research Institute of Police Science, 6-3-1 Kashiwanoha, Kashiwa 277-0882, Japan
[c] Criminal Scientific Investigation Laboratory, Metropolitan Police Department of Tokyo, 3-35-21, Shakujiidai, Nerima, Tokyo 177-0045, Japan

ARTICLE INFO

Article history:
Received 15 April 2015
Received in revised form 14 July 2015
Accepted 15 July 2015
Available online 16 July 2015

Keywords:
Allele dropout probability
Logistic regression
3500xL Genetic Analyzer
Analytical threshold

ABSTRACT

Phenomena called allele dropouts are often observed in crime stain profiles. Allele dropouts are generated because one of a pair of heterozygous alleles is underrepresented by stochastic influences and is indicated by a low peak detection threshold. Therefore, it is important that such risks are statistically evaluated. In recent years, attempts to interpret allele dropout probabilities by logistic regression using the information on peak heights have been reported. However, these previous studies are limited to the use of a human identification kit and fragment analyzer. In the present study, we calculated allele dropout probabilities by logistic regression using contemporary capillary electrophoresis instruments, 3500xL Genetic Analyzer and 3130*xl* Genetic Analyzer with various commercially available human identification kits such as AmpFℓSTR® Identifiler® Plus PCR Amplification Kit. Furthermore, the differences in logistic curves between peak detection thresholds using analytical threshold (AT) and values recommended by the manufacturer were compared. The standard logistic curves for calculating allele dropout probabilities from the peak height of sister alleles were characterized. The present study confirmed that ATs were lower than the values recommended by the manufacturer in human identification kits; therefore, it is possible to reduce allele dropout probabilities and obtain more information using AT as the peak detection threshold.

1. Introduction

DNA profiling using short tandem repeats (STRs) has been extensively applied in human identification and/or kinship analysis. It is possible to obtain a sufficiently high level of discrimination and high likelihood ratios in these types of analyses when a sufficient amount of DNA is available for the analysis and the alleles in the samples can be detected. However, it is often difficult to obtain adequate amounts of DNA to analyze casework samples in criminal investigations, and the stochastic effects such as allele dropout must be considered to carefully interpret the results [1–3]. Although earlier analyses generally eliminate the involvement of a suspect when the suspect's allele was not detected in DNA genotyping performed using casework samples, in recent years, there have been active discussions on how likelihood ratios in mixture interpretation should be calculated in order to consider the possibility of allele dropout [4]. Thus, it is important to evaluate allele dropout probabilities statistically.

Gill et al. recommended to the DNA Commission of International Society of Forensic Genetics that STR typing results be evaluated in terms of allele dropout probabilities [3]. On the other hand, Tvedebrink et al. proposed an estimation model for allele dropout probabilities by logistic regression using AmpFℓSTR® SGM® Plus PCR Amplification Kit (Applied Biosystems) and AmpFℓSTR® SEfiler® Plus PCR Amplification Kit (Applied Biosystems) [5,6], as well as the need to establish a standard for evaluating respective allele dropout probabilities to correctly interpret results when a novel capillary electrophoresis fragment analyzer and other human identification PCR amplification kits (HID Kits) are used for DNA analysis.

3500xL Genetic Analyzer (3500xL) is the latest capillary electrophoresis fragment analyzer, which was released by Applied Biosystems in 2010. Compared with the previous 3130*xl* Genetic Analyzer (3130*xl*), the 3500xL has relative fluorescence units

* Corresponding author at: Department of Forensic Medicine, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo, Tokyo 113-8421, Japan.
E-mail address: s-inoku@juntendo.ac.jp (S. Inokuchi).

http://dx.doi.org/10.1016/j.legalmed.2015.07.006
1344-6223/

(rfu) that are significantly reformed because platforms such as the detection system and optical devices have markedly improved [7,8]. Furthermore, the 3500xL consists of 24 capillaries, which allows forensic laboratories to conduct high-throughput electrophoresis analysis. In this report, primarily, we calculated the analytical threshold (AT), which is the minimum value at which noise can be determined on the basis of the signals, in either the 3500xL or 3130*xl* using three HID Kits [AmpFℓSTR® Identifiler® PCR Amplification Kit (Applied Biosystems) (Identifiler), AmpFℓSTR® Identifiler® Plus PCR Amplification Kit (Applied Biosystems) (IdentifilerPlus), and PowerPlex® 18D System (Promega, USA) (PowerPlex18D)]. Then, logistic curves were generated using SPSS software according to the logistic regression method of Gill et al. [3]. Furthermore, we calculated identical allele dropout probabilities using the values recommended by the manufacturer as the peak detection threshold (175 rfu in 3500xL and 50 rfu in 3130*xl*) [7,9–12] and evaluated the differences in the behavior of the logistic curve.

2. Materials and methods

2.1. Samples and DNA extraction and quantification

DNA profiles using buccal samples from three Japanese volunteers who were predetermined to be heterozygous for 15 loci were generated using Identifiler and IdentifilerPlus (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA). DNA extraction was performed using Qiagen EZ1 Investigator (Qiagen, UK) following the manufacturer's instructions. DNA quantification was performed using Smart Cycler® II (Cepheid, Sunnyvale, CA, USA) and a quantitative PCR method that targeted a human-specific 207-bp DNA sequence present in the D17Z1 region using real-time PCR as previously described [13]. The Ethical Review Board of National Research Institute of Police Science approved the procedure for sample collection.

2.2. PCR amplification

Fifteen loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA) and *amelogenin* were amplified by multiplex PCR using Identifiler and IdentifilerPlus, and 17 loci, including the 15 abovementioned loci along with the two additional loci Penta D and Penta E, were amplified by PowerPlex18D. All reactions were performed, leading to a 25-μL reaction volume, according to the manufacturer's instructions [9–11]. Amplifications were performed using GeneAmp® PCR system 9700 (Applied Biosystems, UK) with thermal profiling following the manufacturer's recommendations [9–11]; PCR amplification consisted of 28 cycles in Identifiler, 28 or 29 cycles in IdentifilerPlus, and 29 cycles in PowerPlex18D.

2.3. Electrophoresis and data analysis

Detection was performed by capillary electrophoresis using the 3500xL with injection parameters set at 1.2 kV for 24 s, whereas the 3130*xl* used an injection parameter of 3 kV for 10 s. Samples were prepared for the 3500xL by adding 1 μL of the PCR product and 0.5 μL of GeneScan™ 600LIZ Size Standard v2.0 (Applied Biosystems, UK) to 8.5 μL of Hi-Di™ formamide (HiDi) (Applied Biosystems, UK); for the 3130*xl*, 1 μL of the PCR product and 0.1 μL of GeneScan™ 500LIZ Size Standard (Applied Biosystems, UK) were added to 8.9 μL of HiDi. For PowerPlex18D, the following conditions were employed: 1 μL of the PCR product, 8.5 μL of HiDi, and 0.5 μL of a CC5 Internal Lane Standard (Promega, USA) were added to each instrument following the manufacturer's instructions [11]. Output data were analyzed using GeneMapper® *ID-X* Software v.1.2 (Applied Biosystems, UK) and GeneMapper® *ID* Software v.3.2.1 (Applied Biosystems, UK), and peak detection was performed using the parameters recommended by the manufacturer [9–11].

2.4. AT analysis

AT values were calculated using a negative control that was amplified only with ultra-pure water in Identifiler, IdentifilerPlus, and PowerPlex18D. The negative control was independently amplified seven times, and the estimated data were used for the injection of the PCR products six times in the 3500xL and four times in the 3130*xl*. The peak detection threshold for the data analysis was 1 rfu for each fluorescent dye, and the partial analysis area ranged from 2,300 data points to 10,000 data points. All analyses were performed using the corresponding panel and bin setting distribution provided by each manufacturer. AT values were calculated according to the following two types of computational methods that utilized peaks of each fluorescent dye recognized by the software using the above conditions. AT values were assessed using the maximum rfu among the obtained values as follows [14,15]:

$$\text{Method 1} = 2 \times (Y_{max} - Y_{min})$$

$$\text{Method 2} = \text{average rfu} + 10 \times \text{standard deviations(SD)}$$

2.5. Allele dropout probabilities by logistic regression

Buccal samples were manually set up using a dilution range with the following DNA template input: 10, 30, 50, and 100 pg with Identifiler, IdentifilerPlus (28 and 29 cycles, respectively), and PowerPlex18D for each in quadruplicate, producing 704 loci (Identifiler and IdentifilerPlus) and 800 loci (PowerPlex18D). Allele dropout was observed using the peak detection threshold set for each AT or the value recommended by the manufacturer. Subsequently, to perform logistic regression, the higher peak height of one of the detected alleles in the locus that did not present any allele dropout or the detected peak height of a locus presented allele dropout was labeled a dummy variable according to the following equation:

(a) Two alleles present = 0.
(b) One allele present and the other absent = 1.

We defined D as the contributor's allele has dropped out, h as the peak height of the present allele. Then, logistic regression was performed using SPSS software and the respective β_0 and β_1 values and allele dropout probabilities [Pr(D)] were estimated using the following equation [3]:

$$\Pr(D) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 h)}} = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 h)\}}$$

$$\text{logit } \Pr(D) = \ln\left[\frac{\Pr(D)}{1 - \Pr(D)}\right] = \beta_0 + \beta_1 h$$

3. Results

3.1. AT analysis

Data on each fluorescent dye were obtained for each HID kit from called alleles recognized using 1 rfu as the peak detection