Forensic Science International: Genetics 20 (2016) 6–19

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Validating multiplexes for use in conjunction with modern interpretation strategies

Duncan Taylor[a,b,*], Jo-Anne Bright[c], Catherine McGoven[c], Christopher Hefford[a], Tim Kalafut[d], 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] Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland 1142 New Zealand
[d] U.S Army Criminal Investigation Laboratory, Defense Forensic Science Centre, Forest Park, GA 30297, USA

ARTICLE INFO

Article history:
Received 22 March 2015
Received in revised form 21 September 2015
Accepted 22 September 2015
Available online 26 September 2015

Keywords:
Forensic DNA interpretation
GlobalFiler
Continuous DNA interpretation
STRmix
Modelling
DNA mixtures

ABSTRACT

In response to requests from the forensic community, commercial companies are generating larger, more sensitive, and more discriminating STR multiplexes. These multiplexes are now applied to a wider range of samples including complex multi-person mixtures. In parallel there is an overdue reappraisal of profile interpretation methodology. Aspects of this reappraisal include

1. The need for a quantitative understanding of allele and stutter peak heights and their variability,
2. An interest in reassessing the utility of smaller peaks below the often used analytical threshold,
3. A need to understand not just the occurrence of peak drop-in but also the height distribution of such peaks, and
4. A need to understand the limitations of the multiplex-interpretation strategy pair implemented.

In this work we present a full scheme for validation of a new multiplex that is suitable for informing modern interpretation practice. We predominantly use GlobalFiler™ as an example multiplex but we suggest that the aspects investigated here are fundamental to introducing any multiplex in the modern interpretation environment.

1. Introduction

In response to requests from the forensic community, commercial companies are generating larger, more sensitive, and more discriminating STR multiplexes. For example in 2010, the CODIS Core Loci Working Group was formed to investigate the expansion of the minimum load criteria to CODIS from 13 STR loci. One of the aims was to balance the total number of loci recommended with the level of discrimination offered in order to reduce the likelihood of adventitious matches and in anticipation of more transnational sharing of DNA profile information [1]. The gender determining locus Amelogenin, 18 autosomal STRs and one Y STR are the new minimum recommended STR marker set with another three autosomal STRs strongly recommended [1,2]. In Europe a core of 15 STRs has been designated as the European Standard Set [3].

These multiplexes are now applied to a wider range of samples including complex multi-person mixtures.

In parallel there is an overdue reappraisal of profile interpretation methodology. Aspects of this reappraisal include

1. The need for a quantitative understanding of allele and stutter peak heights and their variability,
2. An interest in reassessing the utility of smaller peaks below the often used analytical threshold,
3. A need to understand not just the occurrence of peak drop-in but also the height distribution of such peaks, and
4. A need to understand the limitations of the multiplex-interpretation strategy pair implemented.

In this paper we will outline a scheme for the validation of a multiplex that is suitable for use with modern interpretation strategies such as the semi and fully continuous systems being implemented in many parts of the forensic community. We emphasize that it is the multiplex-interpretation method couplet that requires validation. Hence slightly different suggestions might

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

http://dx.doi.org/10.1016/j.fsigen.2015.09.011
1872-4973/

result for the same multiplex with a semi-continuous model than with a continuous model. We differentiate between developmental validation and internal validation. Developmental validation in this paper means actions we suggest should be undertaken by the software developer to ensure that the software is suitable for use on a certain multiplex.

The aspects we suggest should be studied are:

1. Noise,
2. Stutter ratio and variability,
3. Peak height variability both at and between loci,
4. Drop-out,
5. Drop-in, and
6. Saturation of the capillary electrophoresis (CE) camera.

We illustrate this scheme using the GlobalFiler™ and MiniFiler™ multiplexes. The GlobalFiler™ multiplex (Life Technologies, Carlsbad CA) amplifies 22 STRs, the gender marker Amelogenin plus an additional Y-indel locus [4]. The MiniFiler™ multiplex amplifies 8 autosomal STRs plus Amelogenin.

We finish the paper with sections regarding training and general legal acceptance of continuous and probabilistic DNA interpretation systems. While neither of these topics relates directly to kit validation within the laboratory (or at the computer) they remain an important part of any validation and are required before any pairing of software, expert and profiling kit can be introduced and defended in court.

2. Results

2.1. Analytical threshold

The change to probabilistic systems invites a reappraisal of our approach to setting the analytical threshold (AT). This is because modern systems can manage low level peaks better.

Two interpretation strategies are available:

1. A threshold (AT) based approach and
2. Systems that deal with potential noise at the interpretation stage and require no AT.

A peak in the electropherogram (epg) may be allelic, a PCR by-product, artefactual such as pull-up, or electronic noise. Back stutter is almost unavoidable and we can assume that almost every allelic peak has an associated back stutter peak. Forward stutter and double back stutter are also produced by the PCR process, but in smaller amounts. Since back stutter, forward stutter and double back stutter are allelic products they do not differ from a true allelic peak in any way and cannot be differentiated by visual examination. These are not the only artefactual PCR products. For example there is a −2 base pair stutter-like product at SE33 which is a complex locus with largely tetranucleotide repeats.

Discussions of the position of the AT usually concentrate on the electronic noise and it is suggested that the AT should not be used to manage artefacts. For example SWGDAM [5] states:

> . . . the analytical threshold should be established based on signal-to-noise considerations (i.e., distinguishing potential allelic peaks from background). The analytical threshold should not be established for purposes of avoiding artifact labeling as such may result in the potential loss of allelic data.

Valid efforts have been made to model electronic noise and we give a feel of these types of efforts in Appendix 1. These approaches usually consider the probability of a peak of height O_a if it is electronic noise, Pr(O_a|electronic noise). They suggest selecting an AT at some point when Pr(O_a|electronic noise) is expected to be small. We embrace the validity of the sentiment about not using an AT to manage artefacts but in a brutally pragmatic sense it is necessary to consider the downstream effects of the position of the AT. This needs a lot more than a consideration of Pr(O_a|electronic noise) and we suggest that electronic noise is the least difficult of the factors needing consideration. At the time of writing none of the probabilistic systems specifically model forward and double backward stutter. The semi-continuous systems do not model back stutter whereas the continuous ones do.

Consider initially a threshold based approach. Peaks above the AT are often examined manually for morphology. At this stage dye blobs, pull-up and electronic spikes will be removed. Any peak above the AT that passes manual inspection is passed to the interpretation phase. Lowering the AT will detect more allelic peaks but will also pass more artefactual peaks for manual inspection. The total utility of a lowering of the AT is therefore the sum of these effects and depends crucially on how significant the consequences of passing such peaks are. In turn, this depends on how they are treated at the interpretation phase.

In previous binary systems the passing of false peaks had a very significant negative effect (negative utility) on the interpretation. Hence, historically, ATs were set high. The modern systems have a greater resilience to false peaks and hence the utility function is changed. More specifically there is now less risk associated with lower ATs as long as the models within the system for DNA profile behaviors (such as drop-in) are set up accordingly.

The semi-continuous models in widespread use (LRmix [6], Lab Retriever, LikeLTD [7], FST [8] and LiRa [9]) do not utilise peak heights directly in the software. Some interaction of peak height data and semi-continuous systems does exist, for example expert intervention allows for the manual extraction of a clear major [10] and Lab Retriever utilizes peak height in forming the probability of drop-out (D) parameter. All three of these systems currently function with a threshold based strategy and peaks in stutter positions are either removed or dealt with as ambiguous (either partly allelic or totally stutter). Any peaks that are above the AT and passed to the software must be explained as allelic, ambiguous or as drop-in. Peaks dealt with using the drop-in function would include true drop-ins, that is, unreproducible allelic peaks that appear in the profile, and non-allelic peaks not treated as ambiguous. To emphasize that the software should be run by an expert (following Gill and Haned [6]) we will refer to the software-expert pair (SEP). The effect of dropping the AT would be more true allelic peaks detected which has a strongly positive utility. However there will also be more peaks needing manual removal, treatment as ambiguous or drop-in, or which may cause the scientists to artificially increase the assigned number of contributors. These have a negative utility. The net effect is unstudied.

The continuous software programmes in current use (STRmix™ [11] and TrueAllele [12]) can treat noise peaks directly; modelling the probability of these peaks if they arise from a contributor (as allelic or stutter) or if they do not arise from a contributor (encompassing noise or drop-in).

If a drop-in function is used for semi-continuous SEP or STRmix™ the AT value does not need to be set as conservatively as with traditional interpretation methods. The setting of an AT will affect both the probability of drop-out and drop-in. We direct the reader to [13] who explore this idea.

If forward stutter and double back stutter are not manually removed or treated as ambiguous then the drop-in function will now be used to model true drop-ins but also other peaks that pass AT and manual inspection. The drop-in rate will therefore need to be higher than if it was simply modelling drop-in. When used in this way the drop-in rate cannot be set from empirical negative control data but needs to be set from positive samples with known ground truths. In doing this there would be a dependence of profile