Int J Legal Med (2011) 125:87–94
DOI 10.1007/s00414-009-0411-2

TECHNICAL NOTE

Magnitude-dependent variation in peak height balance at heterozygous STR loci

Jason R. Gilder · Keith Inman · William Shields · Dan E. Krane

Received: 16 January 2009 / Accepted: 16 December 2009 / Published online: 10 February 2010

Abstract When the smaller of two peaks at an STR locus is less than 70% the magnitude of the larger peak at that locus, the disparity is typically taken to be an indication that there is more than one contributor of template DNA to the sample being tested. An analysis of 1,763 heterozygous allele pairs suggests that a peak height imbalance threshold that varies with the magnitude of the peaks being evaluated at a locus is superior to a fixed threshold. Identifying samples that are likely to be mixtures and those that are likely to have arisen from a single source is accomplished more reliably when a statistically based, magnitude-dependent peak height imbalance threshold is used. The amelogenin locus was found to behave in a similar fashion and was also found to have no systematic bias that favored the amplification of Y or X alleles.

Keywords Forensic science · DNA typing · DNA mixtures · Peak height imbalance · Peak height ratio · Amelogenin

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0411-2) contains supplementary material, which is available to authorized users.

J. R. Gilder
Forensic Bioinformatics, Inc,
2850 Presidential Drive, Suite 160,
Fairborn, OH 45324, USA

K. Inman
Forensic Analytical Sciences, Inc,
3777 Depot Rd., Suite 409,
Hayward, CA 94545, USA

K. Inman
California State University,
East Bay, 25800 Carlos Bee Boulevard,
Hayward, CA 94542, USA

W. Shields
Department of Environmental & Forest Biology,
College of Environmental Science and Forestry,
State University of New York,
Syracuse, NY 13210, USA

D. E. Krane (✉)
Department of Biological Sciences, Wright State University,
3640 Colonel Glenn Highway,
Dayton, OH 45435-0001, USA
e-mail: Dan.Krane@wright.edu

Introduction

Short-tandem repeat (STR) sequences have become the genetic markers of choice for the purposes of human identification by crime laboratories [1, 2]. STR typing typically involves a polymerase chain reaction (PCR) amplification step followed by size fractionation of the resulting products and fluorescent signal detection and processing. Numerous studies have shown that the signal associated with each allele of a heterozygote at a given locus is approximately equal [3–9]. General practice has shown that the peak height ratio, as measured by dividing the height (in relative fluorescent units (RFUs)) of the lower quantity allele by the height of the higher quantity allele "should be greater than approximately 70% in a single source sample" [3]. When the smaller of two peaks is less than 70% of the height (or, in some instances, the area [10]) of the larger peak at a locus, the disparity in peak heights is typically taken to be an indication that there is more than one contributor of template DNA to the sample. Similarly, if the smaller of two peaks is equal to or greater than 70% of the height (or area) of the larger peak at a locus, the peaks are commonly considered to be consistent

with the proposition that they may have come from a single contributor.

The most comprehensive published study of peak height imbalance is associated with the TWGDAM validation of the AmpFLSTR® PCR amplification kits for forensic DNA casework [7]. That study was based on a relatively small number of allele pairs at heterozygous loci (N=727), considered nine polymorphic STR loci and the amelogenin locus simultaneously, and used data from several different laboratories using platforms with different sensitivities (specifically, the ABI 377 and ABI 310 Genetic Analyzers) [7]. However, the data in that study suggested that peak height imbalance varied in a way that was inversely related to template quantity. Others have also found that peak height imbalance at heterozygous loci becomes exaggerated (even to the point of one allele not being detected at all) when quantities of template DNA are less than those recommended by the manufacturer of commonly used AmpFLSTR® test kits (such as Profiler Plus®, COfiler® and SGM+®) [9, 11]. A constant peak height imbalance threshold such as 70% may incorrectly indicate that a pair of peaks at a locus is from two contributors rather than one when very small amounts of template are available for analysis. Conversely, reliable indications of peak height imbalances due to the presence of more than one contributor to a sample may be missed if a threshold of imbalance suitable for small amounts of template DNA is applied to DNA profiles where larger amounts of template have been amplified. These issues are particularly problematic when interpreting mixed samples.

In this study, we analyzed a large number (N=1,763) of allele pairs at heterozygous loci from reference samples and positive controls that were genotyped by a single laboratory (Forensic Analytical Sciences, Inc.) on a single instrument (an ABI 310 Genetic Analyzer) during the course of routine casework and validation testing. Variability in peak height imbalance relative to DNA quantity (as determined by average peak height of allele pairs at heterozygous loci) was evaluated collectively and on a locus-by-locus basis for 13 commonly used polymorphic STR loci (those amplified by the Profiler Plus® and COfiler® test kits). Trends in the variability of peak height imbalance at the non-STR amelogenin locus in male individuals (where both an X and Y marker are amplified) were also considered separately. In addition, a set of real-world samples were examined to determine how the allele pairs at heterozygous loci were characterized using a magnitude-specific peak height imbalance threshold. The result is a developmental validation of a varying peak height imbalance threshold that is lower for pairs of peaks within an electropherogram that have small average peak heights and higher for pairs of peaks with large average peak heights.

Materials and methods

A total of 1,763 heterozygous STR loci were obtained from 330 sample runs of reference samples and positive controls generated during validation testing (724 allele pairs at heterozygous loci) and routine casework (1,039 allele pairs at heterozygous loci) by Forensic Analytical Sciences, Inc. (Hayward, CA; Online Resource 1, 2). All genotypes were generated in accordance with the laboratory's validated standard protocols (e.g., no additional rounds of amplification were used as might be the case for low-copy-number analyses). A separate dataset was created from 56 amelogenin loci in which both the X and Y alleles were amplified (56 of the tested profiles came from males; Online Resource 3). All DNA profiles were generated with either the Profiler Plus® and/or COfiler® (Applied Biosystems Inc., Foster City, CA) commercial testing kits (294 and 36, respectively) between 1998 and 2008. Each run was performed on an Applied Biosystems 310 Genetic Analyzer. Loci were only considered to be heterozygous if two peaks over 50 RFUs in height were observed and found to be on-ladder by the GeneScan® and Genotyper® software. In those instances where at least one of the detected alleles at a locus was determined to have peak heights in excess of 4,000 relative fluorescent units (RFUs), the locus was not considered further (29 loci were excluded for this reason). When multiple injections of samples were available, only the first injection was used (all first injections were successful).

An average peak height at each heterozygous STR and each male amelogenin locus was calculated by simply dividing the sum of the heights of the two alleles observed at a locus by two. Heterozygous STR loci were sorted by average peak height and separated into 17 bins that each contain 100 heterozygotes (a final 18th bin contains 63 heterozygotes with the highest average peak heights). The average peak height sorted heterozygous STR, and male amelogenin loci were also separated into quintiles.

Linear and logarithmic regression calculations were carried out using Microsoft Excel. Correlation p values were calculated, and t tests were performed using standard statistical formulae [12].

An additional set of 57 allele pairs at heterozygous loci associated with seven known samples prepared to mimic casework samples and tested during the course of the internal validation studies of Forensic Analytical Sciences, Inc. (Hayward, CA) between 2001 and 2008 were evaluated. The samples include four cigarette butts, a licked envelope flap, a hair root, and chewed sunflower seeds. Each heterozygous locus was evaluated using the magnitude-specific peak height imbalance ratio threshold to determine if the peaks were in balance.

Results

The observed range of peak height imbalance ratios was much greater for loci with low average peak heights than those with larger average peak heights (Fig. 1). Peak height ratios were distributed in a folded normal fashion (Fig. 2). A significant relationship (r^2=0.51, p=0.0009) between the extent of peak height balance and average peak height within a locus is more clearly visualized when the data are grouped into bins that each contain 100 allele pairs at heterozygous loci on the basis of their average peak height (Fig. 3). A logarithmic regression reveals an even stronger correlation [y=0.02ln(x)+0.75; r^2=0.80] with the peak height ratio asymptotically approaching a maximum of approximately 0.94. Average peak height ratios within these bins of 100 allele pairs at heterozygous loci range from 0.84 to 0.92 (Fig. 3). Only 37 (2.1%) of the 1,763 allele pairs at heterozygous loci included in this analysis exhibit peak height ratios of less than 0.70. A substantial majority (28 of 37; 75.7%) of the allele pairs at heterozygous loci with peak height ratios of less than 0.70 are associated with the less than one third (492 of 1,763) of the allele pairs at heterozygous loci with average peak heights less than 500 RFUs. The single lowest observed peak height ratio of 0.40 (a pair of peaks of 60 and 149 RFUs observed at the D7S820 locus) had an average peak height of 105 RFUs. All of the bins of 100 allele pairs at heterozygous loci exhibit a maximum peak height ratio of between 0.96 and 1.0.

The number of instances of observed heterozygosity at each of the 13 CODIS loci ranged from 19 to 239 as follows: vWA (n=212), TPOX (n=23), THO1 (n=27), FGA (n=223), D8S1179 (n=148), D7S820 (n=174), D5S818 (n=144), D3S1358 (n=239), D21S11 (n=153), D18S51 (n=217), D16S539 (n=28), D13S317 (n=156), and CSF (n=19). Differences in the number of observed heterozygotes at the individual loci were reflective primarily of the relative number of samples tested with the Profiler Plus® versus the COfiler® test kits, but it was also influenced by the discriminating power of each locus. Each individual locus was generally consistent with the trends seen for the combined dataset (data not shown but available at Online Resource 1, 2). The average peak height ratio observed for each locus falls within 3% of the average of the combined data set and within 5% of all other loci.

The amelogenin locus was evaluated separately. The height of 56 Y alleles ranged from 0.80 to 1.36 (μ=0.997; σ=0.14) relative to their associated X alleles. The relationship between peak height balance and average peak height (Fig. 4) was similar to that observed for the 13 polymorphic STR loci (Fig. 1).

The standard deviation within bins containing 100 allele pairs at heterozygous loci sorted by average peak height is negatively correlated with the average peak height for those bins (Fig. 5). A linear regression yields a significant correlation (r^2=0.58, p=0.0002). As with the peak height ratio analysis, a logarithmic regression yields a stronger correlation (r^2=0.82) as the standard deviations asymptotically approach a minimum peak height ratio standard deviation of 0.04.

Dividing the data into quintiles also reveals a significant difference in peak height balances between allele pairs at heterozygous loci with large average peak heights relative to those with small average peak heights (Table 1). With the exception of the 4th and 5th quintiles (which have distributions with identical average peak height ratios and standard deviations), t tests indicate that the average peak

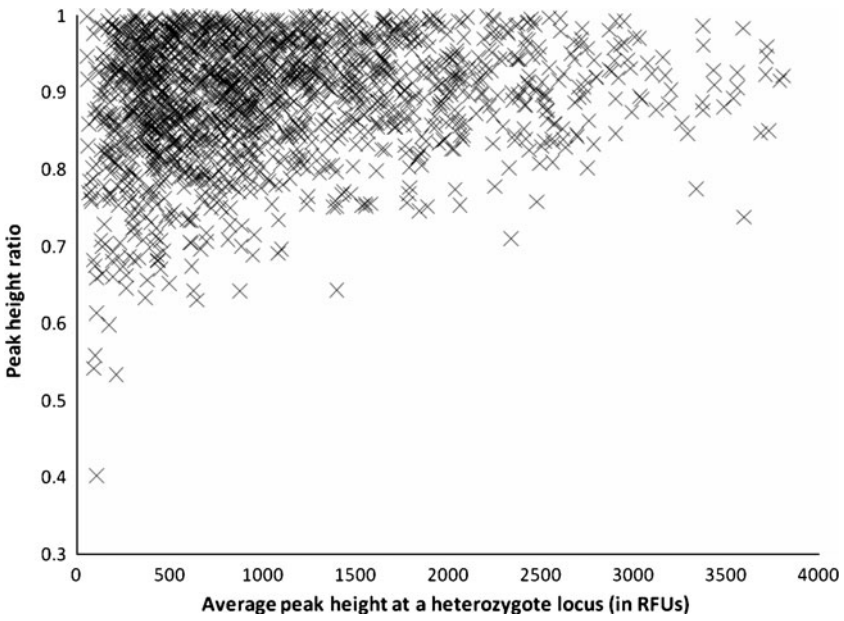

Fig. 1 The average peak height for each heterozygous allele pair and peak height ratio (smaller peak height divided by larger peak height) for 1,763 allele pairs at heterozygous loci (y=2E-5x+0.88; r^2=0.03; p=9.8E-13)

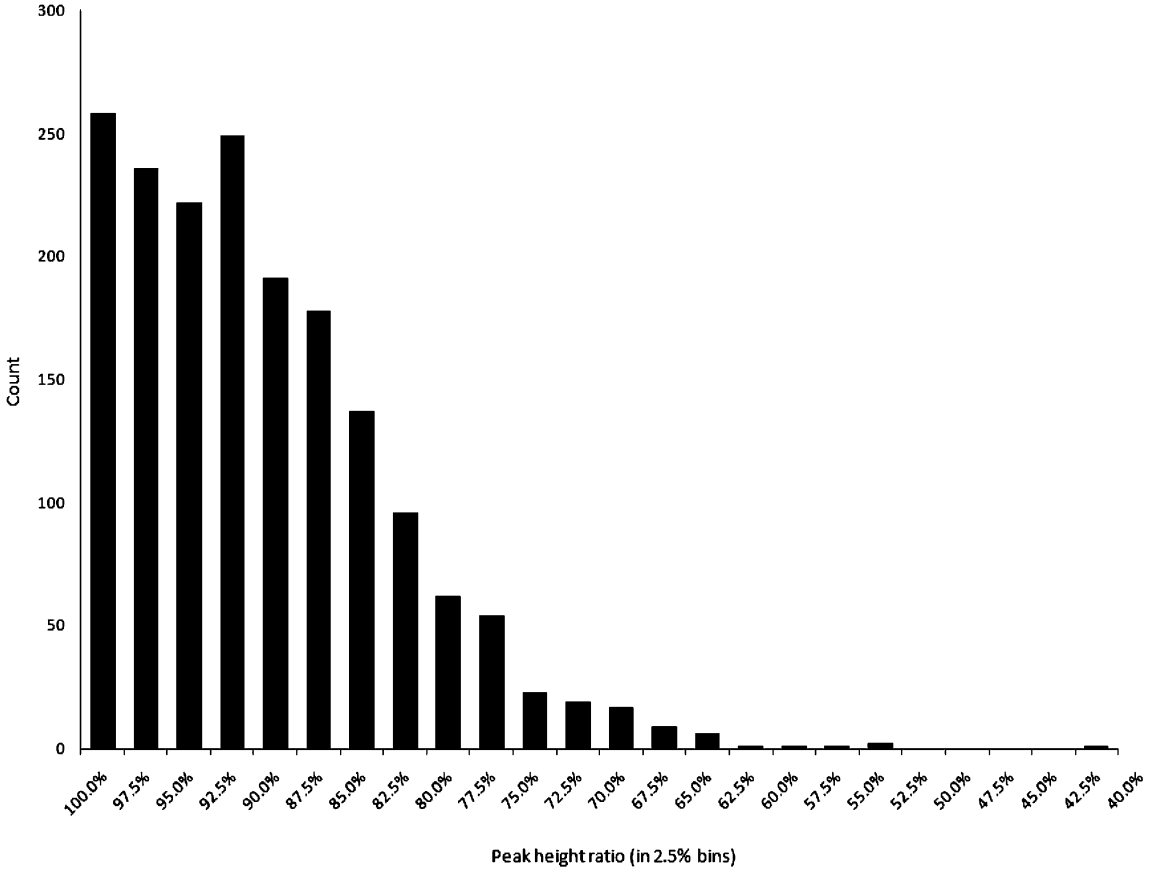

Fig. 2 Histogram of the peak height ratios observed for 1,763 pairs of alleles at heterozygous loci

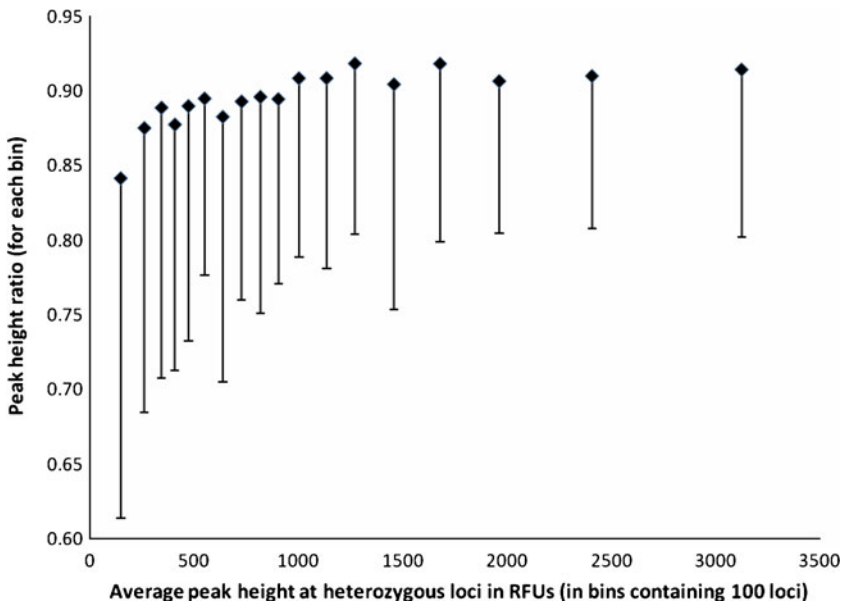

Fig. 3 Average peak height ratio of pairs of alleles at heterozygous loci in RFUs divided into bins containing 100 heterozygotes (sorted by average peak height). The lower 95% decision boundary is shown for each bin (based on the 5th smallest peak height ratio in each bin). A best-fit linear regression line (y=2E-5x+0.88) has a correlation coefficient of 0.51 (p=0.0009) and a best-fit logarithmic regression line [y=0.02ln(x)+0.75] has a correlation coefficient of 0.80. The best-fit logarithmic regression line for the 95% decision boundary [y= 0.059ln(x)+0.36] has a correlation coefficient of 0.81

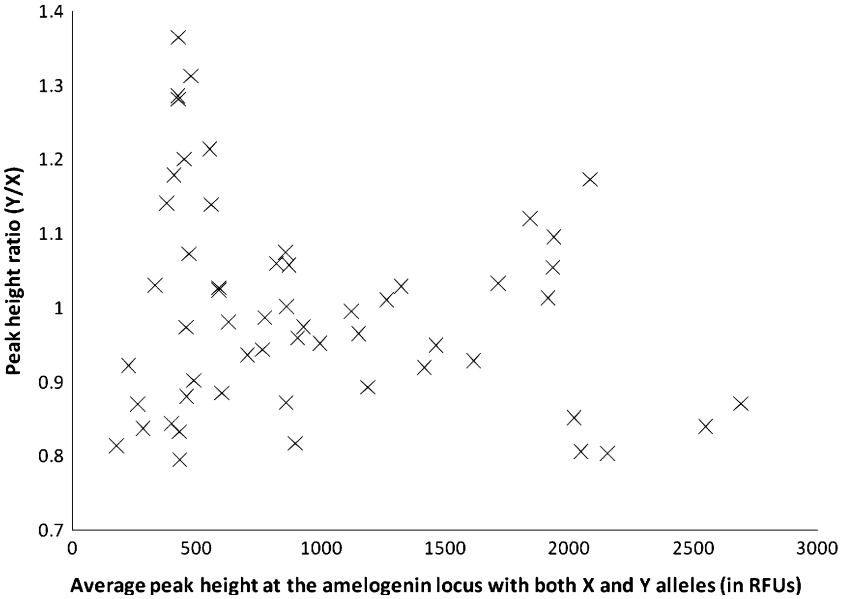

Fig. 4 The average peak height and peak height ratio at the amelogenin locus for 56 samples containing both a Y and X allele. Note that the *x*-coordinate of data points on this plot are determined by dividing the height (in RFUs) of the peak associated with the Y allele by the height (in RFUs) of the X allele (rather than simply dividing the height in RFUs of the smaller of two peaks at a heterozygous locus by the height of the larger of the two peaks as was the case for the plots shown in Fig. 1). A best-fit linear regression for the data was found to be $y=-4E\text{-}05x+1.0379$ with a correlation coefficient of 0.04

height balance ratio for each quintile is significantly different from the others ($p<0.05$; Table 1).

The set of real-world samples shows that almost every heterozygote allele pair is successfully identified using a magnitude-specific peak height imbalance threshold (Fig. 6). There are 55 pairs of alleles out of 57 (96%) that are successfully identified as being in balance using the magnitude-specific peak height imbalance ratio. Only 53 of 57 (93%) would have been similarly identified with a standard 70% threshold. Of the two pairs that were identified as being imbalanced, one was 4% lower than the magnitude-specific threshold (65% vs. 69%), and one was 1% lower than the magnitude-specific threshold (70% vs. 71%). In both cases, the samples exhibited indications of degradation that exacerbated the peak imbalance conditions beyond what would be typically observed.

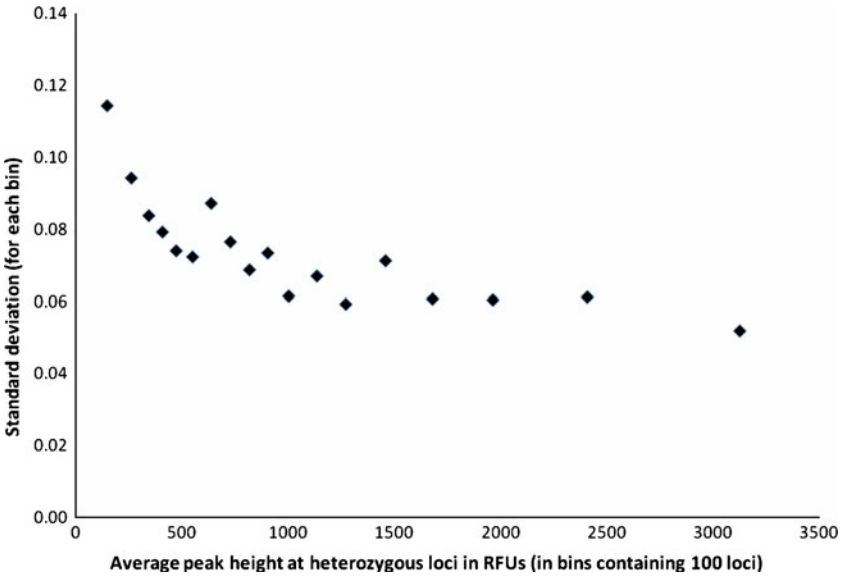

Fig. 5 The average peak height ratio standard deviation divided into bins of 100 heterozygotes (sorted by average peak height). A best-fit linear regression line [$y=-1E\text{-}5x+0.09$] has a correlation coefficient of 0.58 ($p=0.0002$) and a best-fit logarithmic regression line [$y=0.02\ln(x)+0.19$] has a correlation coefficient of 0.82

Table 1 Summary data for 1,763 pairs of alleles at heterozygous loci separated into quintiles (containing 352, 352, 353, 353, and 353 heterozygous loci, respectively) based on average peak height in RFUs

Quintile	Min ave PH	Max ave PH	Ave ave PH	Ave PHR	Min PHR (%)	PHR SD	Fraction of PHR below 70% (%)
1	55	407	271	0.87[a]	40	0.10	6
2	409	683	537	0.88[b]	63	0.08	3
3	683	1,008	843	0.90[c]	64	0.07	1
4	1,010	1,578	1,266	0.91	64	0.06	1
5	1,582	3,808	2,228	0.91	71	0.06	0

PH peak height, *PHR* peak height ratio

[a] Significantly lower (p values ranging from 0.037 to 7.2×10E-11) than all other quintiles based on t tests

[b] Significantly lower (p values ranging from 0.019 to 2.0×10E-7) than quintiles 3, 4, and 5 based on t tests

[c] Significantly lower ($p<0.002$) than quintiles 4 and 5 based on t tests

Discussion

Not all mixtures of two or more individuals exhibit three or more peaks at every locus, particularly when information is only available at a small number of loci such as often occurs with samples containing small amounts of template DNA and/or those that are degraded and/or inhibited [13–21]. In such circumstances, a significant peak height imbalance at a locus where only two alleles are observed may still be indicative of a mixed sample [15–21]. Peak height information is less reliable with low quantities of DNA template mainly due to modifications to the testing process and PCR sampling issues. For example, the most extreme form of peak height imbalance (allelic dropout) is possible even if the paired allele rises above 1,000 RFUs [19] or exhibits an area as high as 9,400 RFUs [14]. Peak height imbalances that exceed a certain threshold in recognized mixtures may also assist in the discernment of a major (and in some circumstances even a minor) contributor's DNA profile. However, the utilization of a *single* peak height ratio threshold for all average peak heights at individual loci is not supported by this evaluation of 1,763 allele pairs at heterozygous loci from 330 reference samples and positive controls genotyped during the course of routine casework and laboratory validation.

The average peak height ratio of known allele pairs at heterozygous loci tends to be lower at loci where the average height of the peaks being considered is small relative to those where they are large (Figs. 1 and 3, Table 1). Conversely, the variation in peak height ratios is greatest at loci with small average peak heights relative to those with large average peak heights. As a result the lower bound of the 95% decision boundary for peak height balance at a heterozygous locus is lower when observed

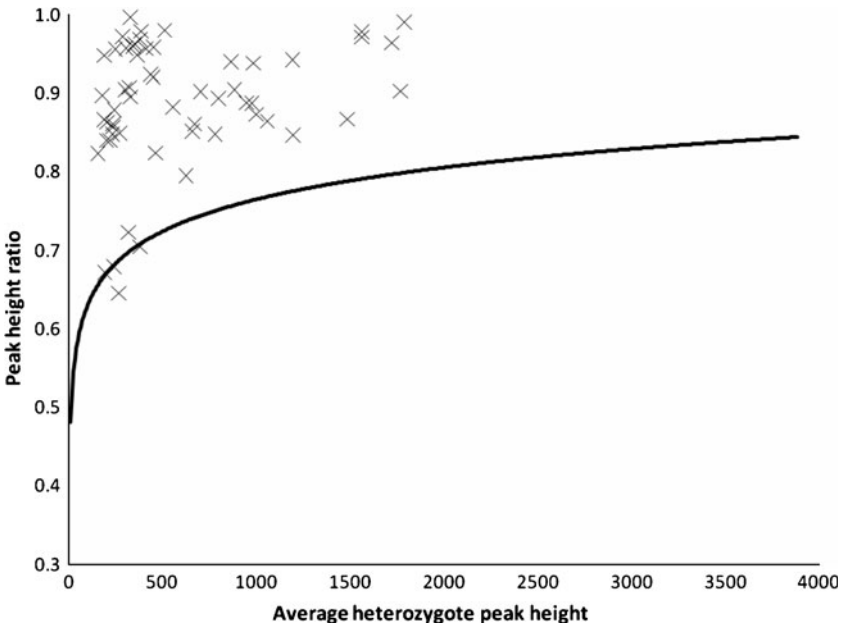

Fig. 6 Plot of the 57 pairs of alleles at heterozygous loci associated with a set of seven real-world samples along with the magnitude-dependent peak height imbalance threshold decision boundary [$y=0.059\ln(x)+0.36$]

peak heights are small relative to when they are large. For instance, this evaluation of the peak height ratios for known heterozygotes suggests that a threshold ratio of 0.61 would be more appropriate for loci where average peak heights are between 55 and 210 RFUs, while a threshold ratio of 0.80 would be more appropriate for average peak heights between 1,560 and 3,808 RFUs (Fig. 3). Using an unnecessarily high threshold ratio for peak height imbalance (e.g., 0.70 for loci with average peak heights between 55 and 210 RFUs) would result in an increased chance of mischaracterizing an unmixed sample as one that is a possible mixture. Similarly, using an unnecessarily low threshold ratio for peak height imbalance (e.g., 0.70 for loci with average peak heights between 1,560 and 3,808 RFUs) would result in an increased chance of mischaracterizing a mixed sample as one that is likely to be unmixed.

The amelogenin locus was also evaluated separately as part of this developmental validation. Despite the fact that differences in the length of the two commonly observed alleles of the amelogenin locus are not due to differences in the number of times that a microsatellite is repeated, the range (0.80 to 1.36) and distribution of observed Y to X peak height ratios (μ=0.997; σ=0.14) was very similar to that of the 13 polymorphic CODIS loci (considering that the first is associated with a two-sided normal distribution and the other is associated with a one-sided normal distribution). The average ratio of the height of Y to X peaks (0.997) suggests that there was no systematic bias that favored the amplification of Y or X amelogenin alleles for these 56 samples.

Many testing laboratories allow their analysts to rely upon their "experience and training" to depart from peak height imbalance thresholds determined during initial validation studies. At loci where two or more alleles are detected, the integration of the statistical lower threshold boundaries derived from the data presented in this study will make the interpretation of DNA samples more reliable and robust. A testing lab can utilize the results from this developmental validation to determine a varying threshold for imbalance based on average RFU values of a pair of alleles observed at a locus (using either the 95% lower confidence boundaries for the binned data available at www.bioforensics.com/phistudy/phistudydata.zip or the best fit logarithmic regression for the overall trend). The 95% logarithmic linear regression decision boundary would identify a pair of peaks as being imbalanced if:

$$\frac{P_1}{P_2} < 0.059 \bullet \ln\left(\frac{P_1 + P_2}{2}\right) + 0.36 \quad P_1 \leq P_2$$

Alternatively, a lab can perform its own peak height imbalance study as part of the validation process of its own testing equipment and protocols.

The use of peak height imbalance thresholds other than the commonly used 70% is not novel (e.g., Moretti et al. use a peak height ratio as low as 59% [8]), and the new versions of commonly used DNA analysis software allow for the incorporation of magnitude-specific peak height thresholds just like those supported by the analysis presented here. For instance, Applied Biosystems GeneMapper® *ID-X* allows for the use of multiple peak height thresholds based on the heights of the observed peaks. Studies such as the one described here provide a systematic, objective, statistically based means of establishing what those magnitude-specific peak height thresholds should be. Future versions of DNA analysis software may be able to bypass the need for establishing multiple thresholds altogether by utilizing a 95% logarithmic linear regression function such as the one derived in this developmental validation study.

References

1. Edwards A, Hammond HA, Lin J, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253
2. Frégeau CJ, Fourney RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. Biotechniques 15:100–119
3. Butler JM (2001) Forensic DNA typing. Academic, San Diego
4. Frank WE, Llewellyn BE, Fish PA et al (2001) Validation of the AmpFLSTR™ Profiler Plus PCR amplification kit for use in forensic casework. J Forensic Sci 46(3):642–646
5. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B (2001) Validation of STR typing by capillary electrophoresis. J Forensic Sci 46(3):661–676
6. LaFountain MJ, Schwartz MB, Svete PA, Walkinshaw MA, Buel E (2001) TWGDAM validation of the AmpFLSTR Profiler Plus and AmpFLSTR COfiler STR multiplex systems using capillary electrophoresis. J Forensic Sci 46(5):1191–1198
7. Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS (2002) TWGDAM validation of AmpFLSTR™ PCR amplification kits for forensic DNA casework. J Forensic Sci 47 (1):66–96
8. Leclair B, Fregeau CJ, Bowen KL, Fourney RM (2004) Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples. J Forensic Sci 49(5):1–13
9. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and Amelogenin: The AmpFlSTR® Identifiler® PCR amplification kit. J Forensic Sci 49(6):1265–1275
10. Gill P, Sparkes R, Pinchin R, Clayton T, Whitaker J, Buckleton J (1998) Interpreting simple STR mixtures using allele peak areas. Forensic Sci Int 9l:4l–53l
11. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. J Forensic Sci 112:17–40
12. Spiegel MR, Stephens LJ (1999) Schaum's outline on theory and practice of statistics, 3rd edn. McGraw-Hill, New York
13. Budowle B, Hobson DL, Smerick JB, Smith JAL (2001) Low copy number—consideration and caution. Twelfth International

Symposium on Human Identification. Promega Corporation, Madison, Wisconsin, available at http://www.promega.com/GENETICIDPROC/ussymp12proc/contents/budowle.pdf
14. Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis. Forensic Sci Int 123:215–223
15. Schneider PM, Bender K, Mayr WE et al (2004) STR analysis of artificially degraded DNA—results of a collaborative European exercise. Forensic Sci Int 139(2):123–134
16. Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. J Forensic Sci 50:1361–1366
17. Fukshansky N, Bär W (2005) DNA mixtures: biostatistics for mixed stains with haplotypic genetic markers. Int J Legal Med 119:285–290
18. Budowle B, Eisenberg AJ, van Daal A (2009) Validity of low copy number typing and applications to forensic science. Croat Med J 50:207–217
19. Caragine T, Mikulasovich R, Tamariz J, Bajda E, Sebestyen J, Baum H, Prinz M (2009) Validation of testing and interpretation protocols for low template DNA samples using AmpFlSTR® Identifiler®. Croat Med J 50:250–267
20. Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. Int J Legal Med 123: 1–5
21. Gilder J, Koppl R, Kornfield I, Krane D, Mueller L, Thompson WC (2009) Comments on the review of low copy number testing (letter). Int J Legal Med 123(6):535–536