J Forensic Sci, Sept. 2004, Vol. 49, No. 5
Paper ID JFS2003312
Available online at: www.astm.org

Benoît Leclair,[1,2] *Ph.D.; Chantal J. Frégeau,*[1] *Ph.D.; Kathy L. Bowen,*[1] *B.Sc.; and Ron M. Fourney,*[1] *Ph.D.*

Systematic Analysis of Stutter Percentages and Allele Peak Height and Peak Area Ratios at Heterozygous STR Loci for Forensic Casework and Database Samples*

ABSTRACT: To assist the interpretation of STR DNA typing results from forensic casework samples containing mixtures, the range of heterozygous allele peak height and peak area ratios (HR) and stutter percentages (stutter %) for the loci comprised in the AmpFℓSTR® Profiler Plus™ (PP) kit were assessed on 468 database and 275 casework single source samples. Stutter % medians were similar for database and casework samples, ranging from 2% to 7%. The upper limit of the stutter value range was 16%, calculated as median +3 SD, although lower locus-specific values could be used. HR medians were $93 \pm 6.5\%$ for database samples, $88 \pm 12\%$ for casework samples. For casework samples, the maximum signal imbalance noted was 52%, calculated as median −3 SD. No significant difference was observed between peak height and peak area calculated values. This study shows the importance of selecting the proper reference database for the establishment of HR threshold values.

KEYWORDS: forensic science, casework, STR, stutter, heterozygous allele ratio, interpretation

The field of human identification has progressed significantly in recent years with the development of highly discriminating PCR-based DNA typing systems. The use of STR loci and highly sensitive instrumentation based on fluorescence detection enables rapid genetic profiling from minute amounts of biological material. As such, STR DNA typing analysis has been extensively used in forensic casework to establish the presence of a perpetrator's DNA in crime scene evidence. Biological evidence recovered at crime scenes often consists of mixtures of bodily substances originating from more than one individual, so mixed STR profiles are frequently encountered in criminal investigations. Profile interpretation in such cases normally follows a systematic approach (1–3). It involves the identification of potential major and minor components, the assignment, when possible, of each allele peak to a specific contributor, and the identification of shared alleles. If a profile other than that of the victim can be developed, then it can be compared with that of a suspect and/or queried against a criminal intelligence DNA database.

Three major characteristics of the STR DNA typing process can impact on mixture interpretation: 1. Stutter peaks, which are minor peaks typically one repeat unit length shorter than that of their parent nominal allele peak and generated through strand slippage during amplification (4), can interfere with mixture interpretation. As an example, an allele from a minor profile can potentially escape detection if it co-localizes on a chromatogram with the stutter peak of a nominal allele from the major profile. 2. Sporadic imbalances in allele peak height or peak area ratios at heterozygous loci can also impact on mixture interpretation. In theory, when alleles at a heterozygous locus are stochiometrically amplified, allele peak height and area ratios should be close to 1. However, imbalances can be triggered by stochastic effects when amplifying too little DNA or degraded DNA, or by the inability to remove PCR inhibitors during DNA extraction. These allelic imbalances can complicate the interpretation of mixed profiles if allele peaks from one contributor cannot be reliably identified because of significant signal imbalances. 3. The unpredictability concerning the number of shared alleles between contributors can also reduce the ability to identify individual profiles in a mixture.

The interpretation of mixed DNA typing profiles is facilitated by statistics compiled for the range of stutter % and HR values to be expected from single source samples. Such data can be readily collected from profiles derived from population database samples (5–8). However, this sample type is more akin to felon data banking samples, and mixture interpretation guidelines based on data obtained from this sample type may not necessarily be the best possible choice for casework samples. Data on simulated or limited size actual casework sample sets have also been published (9,10). Admittedly, it can prove difficult to assemble, without selection bias, a small-sized representative sampling of casework specimens for such a study as casework samples can be exposed to environmental insult(s) varying in nature, intensity and duration before being recovered from crime scenes. Data from a large number of unselected actual casework single source samples would likely provide more realistic guidelines for the interpretation of mixed profiles encountered in casework specimens. This study, part of a larger validation effort (11), reports the systematic stutter % and HR analysis performed on 275 casework samples processed with the AmpFℓSTR® Profiler Plus™ (PP) kit and the ABD 377 DNA sequencer. A similarly large number of pristine database samples ($n = 468$) was used to provide adequate benchmark values against which casework data could be compared.

[1] National DNA Data Bank, Royal Canadian Mounted Police, 1200 Vanier Parkway, Ottawa, Ontario, Canada, K1G 3M8.

[2] Current address: Myriad Genetic Laboratories, Inc., 320 Wakara Way, Salt Lake City, Utah.

* Work presented at the following meeting: Profiler Plus Validation: Establishing thresholds of stutter percentages and allele ratio at heterozygous loci for use in mixture analysis. Ninth International Symposium on Human Identification, Orlando, Florida, 1998.

Received 13 Sept. 2003; and in revised form 24 April 2004; accepted 24 April 2004; published 4 Aug. 2004.

Copyright © 2004 by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.

Materials and Methods

Biological Samples

A total of 164 questioned samples and 111 known samples from 48 adjudicated casework files were used as our casework specimens. A total of 468 samples from various anonymous population databases were used as our database sampling. All samples were extracted according to standard operational protocols (one-step organic extraction followed by Microcon-100 concentration (12)) and quantitated using the ACES 2.0 chemiluminescence kit (Whatman, Clifton, NJ).

PCR Amplification, Pristine Sample Mixtures

Samples (2.5 ng) in this study were amplified with the AmpFℓSTR® Profiler Plus™ kit (Applied Biosystems, Foster City, CA), in 25 μL reaction volumes, and strip-capped thin walled 0.2 mL Perkin Elmer MicroAmp™ reaction tubes. The cycling conditions were the following: 95°C–11 min, once; 94°C–60 s, 59°C–90 s, 72°C–90 s, for 28 cycles; 60°C–30 min, once; 22°C–overnight (13,14). All amplifications were carried out in Perkin-Elmer Gene Amp™ 9600 DNA Thermal Cyclers.

Electrophoresis

Profiles were resolved on ABD 377 DNA Sequencers. A 1.5 μL aliquot of each amplified PCR reaction was diluted in 4.5 μL of loading buffer (2X Tris-Borate-EDTA, 20 mM EDTA, 20 mg/mL blue dextran, 0.5 μL GS 500 ROX (ABD), 9 M urea), heat-denatured at 95°C for 2 min and snap-cooled on ice-water. With allelic ladders, a volume of 0.7 μL of each of the 3 ladders provided with the PP kit were mixed together with the above-mentioned volumes of loading buffer. For both amplicons and allelic ladders, a volume of 1.5 μL from these mixes was then loaded onto 0.2 mm thick, 4% 19:1 acrylamide (Bio-Rad): bis-acrylamide (BRL) 6 M urea gels. All gels were cast on 36 cm WTR plates with square-tooth combs; the gels had been polymerized by making the solution 0.05% for both TEMED and APS, cured for 2 h and pre-run for 30 min at 1000 V. Electrophoresis was carried out at 3000 V for 2.5 h at 51°C in 1X TBE. On each gel, two lanes were reserved for allelic ladder samples.

Data Processing

Sample data were analyzed with the GeneScan Analysis software (v. 2.1) under the Local Southern sizing algorithm. Allele size and designation as well as peak height and peak area values for every allele in the profiles were exported from Genotyper 2.0 into a spreadsheet for statistical calculations. Nominal allele peaks under 100 relative fluorescence units of peak height intensity, as well as casework samples exhibiting any signs of an additional minor profile were excluded from both the stutter % and HR tabulations. For heterozygous loci showing alleles differing by one repeat unit, the largest allele was excluded from stutter % calculations as the stutter peak of the largest allele would overlap with the peak of the smallest allele. Stutter % and HR were calculated from peak height and peak area to document any advantages the use of one type of data might have over the other.

HR ratios were calculated in two ways: as peak height or area of the weaker intensity allele peak over that of the stronger intensity allele peak (referred to herein as Method #1), the most widely used method; as peak height or area of the longer allele over that of the shorter allele (referred to herein as Method #2).

Results and Discussion

STR DNA typing profiles derived from forensic casework samples often present evidence of more than one contributor. It is often possible to dissect out, to some extent, the contributing genotypes from such mixtures. The process of assigning source attribution of any given allele in the STR profile of a mixture is generally complicated by the fact that mixtures may include more than two contributors, that each contributor may prove to be homozygous or heterozygous at the loci being examined, and that a variable number of alleles may be shared. In that regard, mixed profile interpretation would be facilitated in the absence of stutter (as these peaks can co-localize with those of a minor contributor), and if alleles within a locus were always perfectly balanced for peak height or peak area. However, stutter is a naturally occurring DNA polymerase extension artifact that is believed to have caused STRs to appear in nature. Uneven peak height ratios can be caused by primer binding site polymorphisms: even with the best primer design, primer binding site polymorphisms may still be encountered when a large enough collection of specimens is genotyped at a given number of STR loci. Uneven peak height ratios can also be caused by amplification under stochastic conditions, and such conditions are regularly encountered with forensic specimens. Given these operational limitations, guidelines developed from empirical data can assist in the interpretation of mixed profiles, and allow for artifacts to be distinguished from the presence of a second contributor to a mixture. It is generally accepted that, for shared alleles in a mixed profile, peaks exceeding the range of stutter % and HR (in height or area) established from samples reputed to be of single source should be considered significant indication of the possible presence of a mixture (5–10). Such a tentative conclusion based on the peak pattern observed at a single STR locus may find further substantiation in the examination of peak patterns at additional STR loci. Several studies have reported stutter % and HR ranges obtained from population database, mock sample testing or operational casework DNA typing data. It is understood that casework samples can be subjected to combinations of environmental insults varying in nature, intensity and duration. Simulated casework samples subjected to a single individual environmental insult represent good approximations of reality but may not truly reflect the complex nature of actual casework samples. The aim of this study was to provide stutter % and HR ranges derived from a large collection of casework samples processed with the PP kit. In order to obtain a measure of the increase in range that can be attributed to the sub-optimal conditions of casework samples, a similarly large number of population database samples were also processed to provide an adequate benchmark for comparison.

Stutter %

The strand slippage model proposes that the stutter phenomenon is caused by slippage that takes place during pauses of the DNA polymerase complex during primer extension as a consequence of the enzyme's limited processivity and the presence of secondary structures within repeated sequences (4,15–19). As the core repeat sequences of all nine loci of the PP kit have a 25% GC content (see Table 1), which level of GC content favors strand dissociation of double-stranded DNA, it is reasonable to postulate that, during an extension pause of the DNA polymerase complex anywhere within the STR repeat cluster, the strand dissociation required for a subsequent out-of-register re-annealing event would be facilitated. Whether the two strands re-anneal with their original complement or with an out of register position is likely

TABLE 1—*Stutter percentage for population database and casework samples*.

				Population Database									Casework								
				Height				Area					Height				Area				
STR Locus	Dye	Repeat Structure*	Range (Common Alleles)	Median	SD	Median +2 SD	Median +3 SD	Median	SD	Median +2 SD	Median +3 SD	*n*	Median	SD	Median +2 SD	Median +3 SD	Median	SD	Median +2 SD	Median +3 SD	*n*
D3S1358	FAM	TCTA-TCTG	12–19	6.2	1.2	8.6	9.7	6.0	1.3	8.5	9.8	500	6.0	2.3	10.6	12.9	5.5	2.3	10.1	12.4	262
vWA	FAM	TCTA-TCTG	11–21	6.7	1.5	9.8	11.3	6.4	1.8	10.0	11.8	545	5.9	2.6	11.0	13.5	4.9	2.4	9.8	12.3	276
FGA	FAM	TTTC-TTTT	18–30	6.7	2.1	10.9	13.0	6.6	2.1	10.8	12.9	687	5.2	2.6	10.4	13.0	4.9	2.5	9.9	12.3	296
D8S1179	JOE	TCTA	8–19	5.4	2.6	10.6	13.2	5.0	2.8	10.5	13.3	623	4.7	2.6	9.9	12.5	4.1	2.9	9.9	12.7	296
D21S11	JOE	TCTA	24.2–38	6.3	2.1	10.4	12.5	6.1	2.1	10.3	12.4	724	5.2	2.7	10.5	13.2	4.7	2.9	10.6	13.5	276
D18S51	JOE	AGAA	9–26	7.0	3.1	13.2	16.2	6.8	3.1	13.0	16.1	754	6.0	3.2	12.4	15.6	5.9	3.4	12.6	15.9	325
D5S818	NED	AGAT	7–16	4.1	2.3	8.7	11.0	3.8	2.3	8.4	10.8	713	3.9	2.1	8.1	10.3	3.5	2.1	7.7	9.9	282
D13S317	NED	TATC	8–15	3.4	2.4	8.2	10.5	3.1	2.3	7.7	9.9	707	3.0	2.1	7.3	9.4	2.7	2.1	6.9	9.0	301
D7S820	NED	GATA	6–15	3.6	2.7	9.0	11.7	3.2	2.5	8.2	10.7	650	2.5	2.2	6.9	9.1	2.0	1.9	5.7	7.6	307

SD = standard deviation.
* = from STRBase web site (www.tstl.nist.gov/div831/strbase/).

influenced by strand flexibility to accommodate the formation of the required loop. Under that postulate, the probability of occurrence of a slippage event should increase proportionally with repeat cluster length instead of amplicon length, as polymerase complex pauses outside of the repeat cluster would not be expected to produce stutter.

Median stutter and standard deviation values were calculated for peak height and area data for all nine loci included in the PP kit, and are presented for both population database and casework samples in Table 1. As graphical representations of any dataset can reveal data point distributions that are not adequately reflected in a statistical compilation, the datasets are also presented in Fig. 1 as stutter % plotted against peak height and area values of the parent nominal allele. The stippled area in each individual panel represents the average stutter % $\pm$ 3 SD.

The graphical representation was selected to evaluate data point distributions but also to assess whether extreme fluctuations of peak height and area values of parent nominal alleles have any impact on associated stutter % values. In that respect, although there are fewer data points at higher peak height and area values, the data from Fig. 1 suggests that for all tested loci and sample types, elevated peak height and area values are not associated with higher stutter % values, which suggests a wide linear detection range for the CCD camera of the ABD 377. For peak heights of parent nominal alleles under 500 RFU in value, stutter % values gradually increase with decreasing peak heights of parent nominal allele for many loci, as highlighted in the D18S51 and D5S818 (database) panels of Fig. 1. In these situations, the increase in stutter % can be explained by a diminishing contribution of the true stutter component while the background noise component remains constant in the total stutter peak signal.

As shown in Table 1, median stutter values were lower for casework than for database samples across all loci, for both peak height and peak area data. The presentation format selected in Fig. 1 to display the datasets clearly demonstrates that the observed difference in median values is not a consequence of the presence of higher signal strength off-scale peaks within the population database dataset. The presence of such peaks would cause an underestimation of their true height and area, and an overestimation of their stutter %. A significant difference between population database and casework samples resides in the number of alleles displaying no measurable stutter at D3S1358 and vWA. This phenomenon likely contributes to lowering median values for these loci; however, as the difference is not observed at every locus, it does not entirely explain the overall drop in median for casework samples. NED-labeled loci, featuring uniformly low average number of repeats within their common allele range, display median values significantly lower than their FAM- and JOE-labeled counterparts. This is consistent with data reported by Holt et al. (10), but contrary to data reported by Moretti et al. (9). For casework samples, the median actually drops from D5S818 to D7S820 for NED-labeled loci, despite the increase in amplicon length. However, graphical representation of the same data in Fig. 1 shows bimodal data point distributions as evidenced by large collections of data points on the X axis of graphs for the three NED-labeled loci. Undoubtedly a desirable feature, these "zero" stutter % values displace medians towards lower values at the expense of an increase in the upper limits of the $\pm$3 SD ranges. This suggests that for D5S818, D13S317, and D7S820, median or average stutter values and their associated standard deviations represent very conservative estimates of true stutter % values.

Core STR repeat sequence commonality can be observed in Table 1. D3S1358, vWA, FGA, D8S1179, and D21S11 all share the same sequence and, similarly, share higher stutter % values. Conversely, there is an absence of sequence commonality within the NED-labeled loci. These data suggest that the core repeat base sequence composition may indeed impact on the flexibility of a dissociated strand and its access to an out-of-register position on a complementary strand.

Under the tested hypothesis, D21S11, with 38 repeats for its largest allele on the PP allelic ladder, should boast the largest median value for stutter. With the tested dataset, the D18S51 locus produced the largest median and SD of all loci. A breakdown of stutter % values for each allele for these two loci is presented in Fig. 2. It is clear that, within a locus, as the repeat cluster grows in length, the average stutter value increases, as observed in other studies (5,10). A case could be made that allele specific stutter % thresholds should be used, but for uncommon alleles, too few data points are available to establish statistically relevant median and SD values. Although there are three "X.2" D18S51 variants in the allelic ladder data displayed in Fig. 2, no D18S51 variants were encountered within the tested population dataset, whereas variants are frequent with D21S11 typing. The "X.2" variants of D21S11 display lower stutter values, as reported previously (10), the negative difference likely reflecting the destabilization caused by the attempt at re-annealing out of register the repeat containing the variant with a non-complementary non-variant sequence. The frequent encounter of one or two base pair variants thus contributes to a decrease in the stutter % range for an affected locus. However, the D18S51 data cannot rule out overall amplicon length as a valid cause for larger medians. The D18S51 amplicon length range extends significantly higher than any other loci, which produces generally lower intensity amplicons, as can be seen in Fig. 1. This increases the number of data points under 500 RFU, a range of intensity demonstrated above as generating artificially elevated stutter % values. D18S51 is also unique among PP loci as it has a unique core repeat sequence as well as a unique allele and amplicon length ranges, which make it difficult to ascribe its overall behavior to any of the specific mechanisms described above.

In summary, the stutter % medians varied from locus to locus, increasing with the number of core repeats. D5S818, D13S317 and D7S820 stutter % medians were lower than that of other loci for both database and casework samples. Stutter % medians ranged from 3.4% to 7.0% for peak height, 3.1% to 6.8% for peak area, for population database samples, and from 2.5% to 6.0% for peak height, 2.0% to 5.9% for peak area, for casework samples. Stutter % ranges, defined as the median +2 SD or +3 SD, were similar between the two sample datasets. For mixed profile interpretation, an upper conservative threshold value of 16% (upper range for D18S51 for the median +3 SD) could be used as an overall threshold beyond which a peak in a stutter position should be considered as potentially originating partly from a minor profile. However, a case can be made for locus-specific upper ranges to be used. These ranges should be considered as guidelines, and only once the analysis of the data from all loci is completed can an assessment be made of whether or not a sample reflects the presence of a mixture.

HRs

With optimal amounts of pristine template DNA, a typical PCR reaction contains, for any given locus, an identical number of copies of both allelic targets. In theory, under those conditions, stochiometric amplification should occur and identical amplicon yields from each allelic target should be obtained. In that respect, the PP

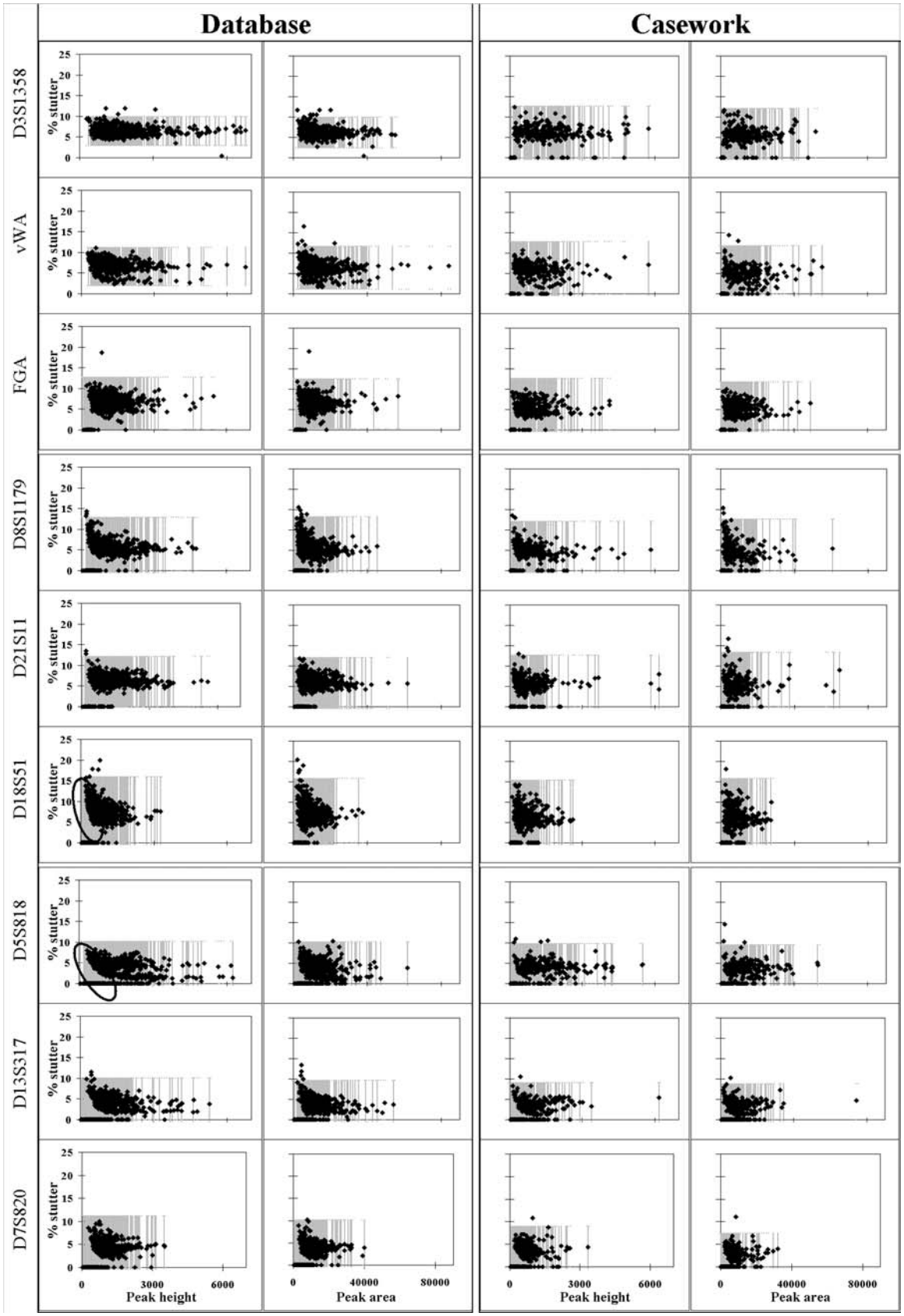

FIG. 1—*Percentage stutter for database and casework samples based on peak height and peak area measurement. Circled data points in peak height database data for D18S51 and D5S818 exemplify the increase in stutter percentage encountered at low peak height and area values. Stippled areas represent the range of percentage stutter covered by the average stutter value ± 3 SD for the entire locus.*

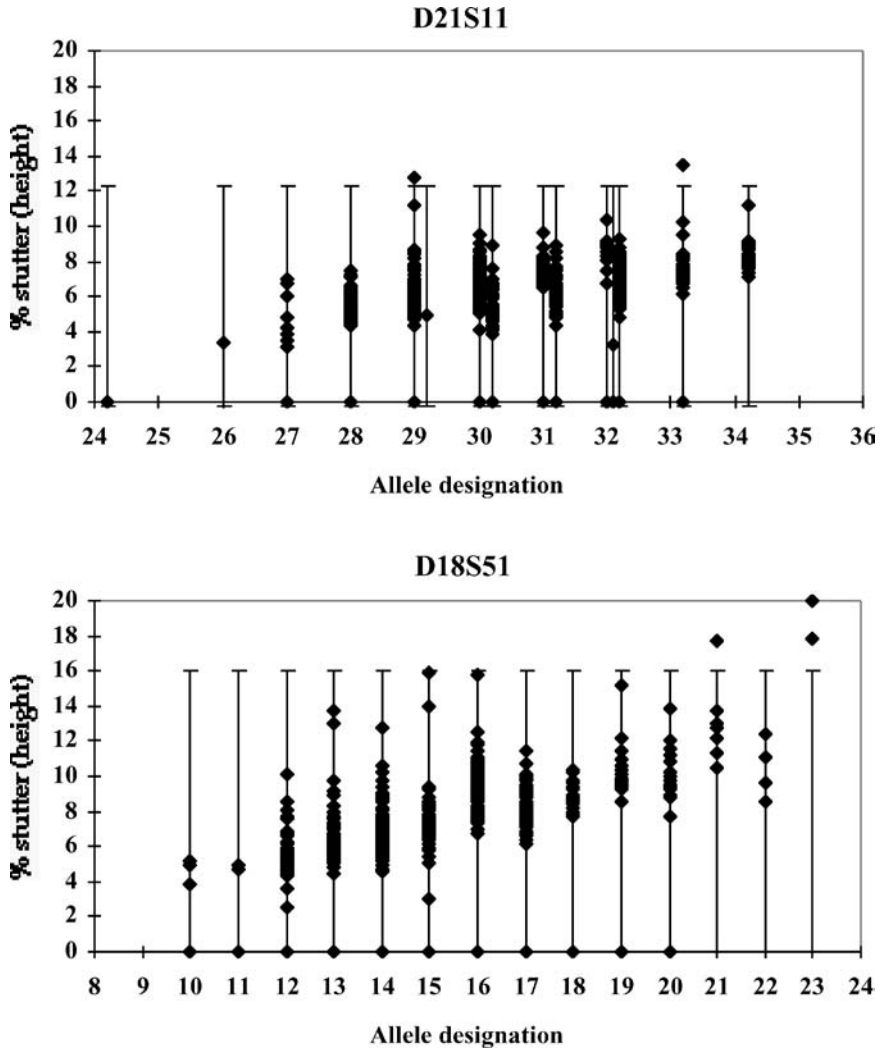

FIG. 2—*D18S51 and D21S11 percentage stutter for database samples. Vertical lines represent the range of percentage stutter covered by the average stutter value ± 3 SD for the entire locus. Off-centered vertical lines and data points at alleles 29, 30, 31 and 34 represent X.2 variants. Off-centered vertical lines and data points at allele 33 represent from left to right, 33.1 and 33.2 alleles, respectively.*

kit and ABD 377 sequencer analytical platform has been shown to be a robust platform for the accurate measurement of profile components (14). However, in practice, because of reactant depletion and accumulation of pyrophosphate as a by-product of the primer extension process in later cycles of the PCR reaction, among other contributing factors, primer extension is believed to be less efficient in the last cycles of amplification (20–24). Amplicons that fail to be completely extended by the end of an extension segment are no longer amplification targets. As a consequence, late into cycling, shorter amplicons are presumed to stand a better chance of getting fully extended than their longer counterparts. Therefore, even under optimal DNA template and PCR conditions, the average ratio of peak height and area of one peak over the other in any allelic pair at any given STR heterozygous locus is anticipated to be slightly less than 1. HR data can provide a useful estimate of the anticipated data point dispersion about an average or median value. HR ratio calculations can be performed in two ways: as peak height and area of the weaker intensity allele peak over that of the stronger intensity allele peak (Method #1), the most widely used method; and as peak height and area of the longer allele over that of the shorter allele (Method #2). If it is to be assumed that the longer allele peak should generally be of less intensity than the shorter allele peak, then both calculation methods will produce the same HR value. However, in practice, this is often not the case. Therefore, both calculation schemes have been used in this study to document any advantages, or lack thereof, of one scheme over the other.

Median HR and SD values were calculated from peak height and area data for all nine loci included in the PP kit, and are presented for both database and casework samples in Table 2. Again, the same dataset is presented in Fig. 3 (Method #1) and Fig. 4 (Method #2) as HR values plotted against peak height and area values of the parent nominal allele. As for stutter % data, this graphical representation was selected to evaluate whether extreme fluctuations of peak height and area values of parent nominal alleles have any impact on associated HR values. The stippled areas in Figs. 3 and 4 represent the average HR value ± 3 SD. In comparison to the population database statistics, SDs were doubled for casework samples for both peak height and peak area, a reflection of the challenge these samples collectively represent for this PCR-based assay.

Both tabular and graphical representations of the data show a much wider data spread for Method #2 when compared to Method #1. Discrete examples of how the two HR calculation schemes reflect actual allele peak height and area imbalances are shown

TABLE 2—*Heterozygous peak height/area ratios for population database and casework samples calculated using two different methods.*

	Method #1: Het Ratio (Lowest Intensity Allele/Highest Intensity Allele)										Method #2: Het Ratio (Longer Allele/Shorter Allele)									
	Database					Casework					Database					Casework				
	Height		Area			Height		Area			Height		Area			Height		Area		
STR Locus	Median	SD	Median	SD	*n*	Median	SD	Median	SD	*n*	Median	SD	Median	SD	*n*	Median	SD	Median	SD	*n*
D3S1358	91%	6%	93%	6%	638	88%	12%	90%	13%	174	92%	9%	94%	10%	638	90%	19%	91%	23%	174
vWA	92%	7%	92%	7%	714	85%	12%	88%	11%	160	94%	11%	95%	12%	714	87%	17%	92%	19%	160
FGA	92%	6%	92%	6%	802	88%	12%	88%	12%	163	94%	9%	93%	10%	802	93%	22%	92%	23%	163
D8S1179	92%	6%	94%	6%	718	89%	13%	89%	14%	161	93%	8%	97%	10%	718	92%	17%	95%	20%	161
D21S11	94%	6%	93%	6%	792	90%	11%	89%	13%	173	96%	9%	96%	10%	792	95%	20%	94%	25%	173
D18S51	91%	7%	92%	7%	782	84%	12%	87%	12%	177	93%	12%	94%	11%	782	88%	21%	91%	20%	177
D5S818	91%	7%	94%	5%	710	89%	12%	90%	12%	154	91%	9%	96%	8%	710	90%	16%	92%	17%	154
D13S317	92%	7%	93%	6%	728	90%	10%	90%	11%	150	95%	11%	95%	11%	728	94%	19%	93%	23%	150
D7S820	91%	8%	93%	6%	656	88%	13%	90%	14%	157	94%	12%	95%	10%	656	91%	18%	92%	19%	157

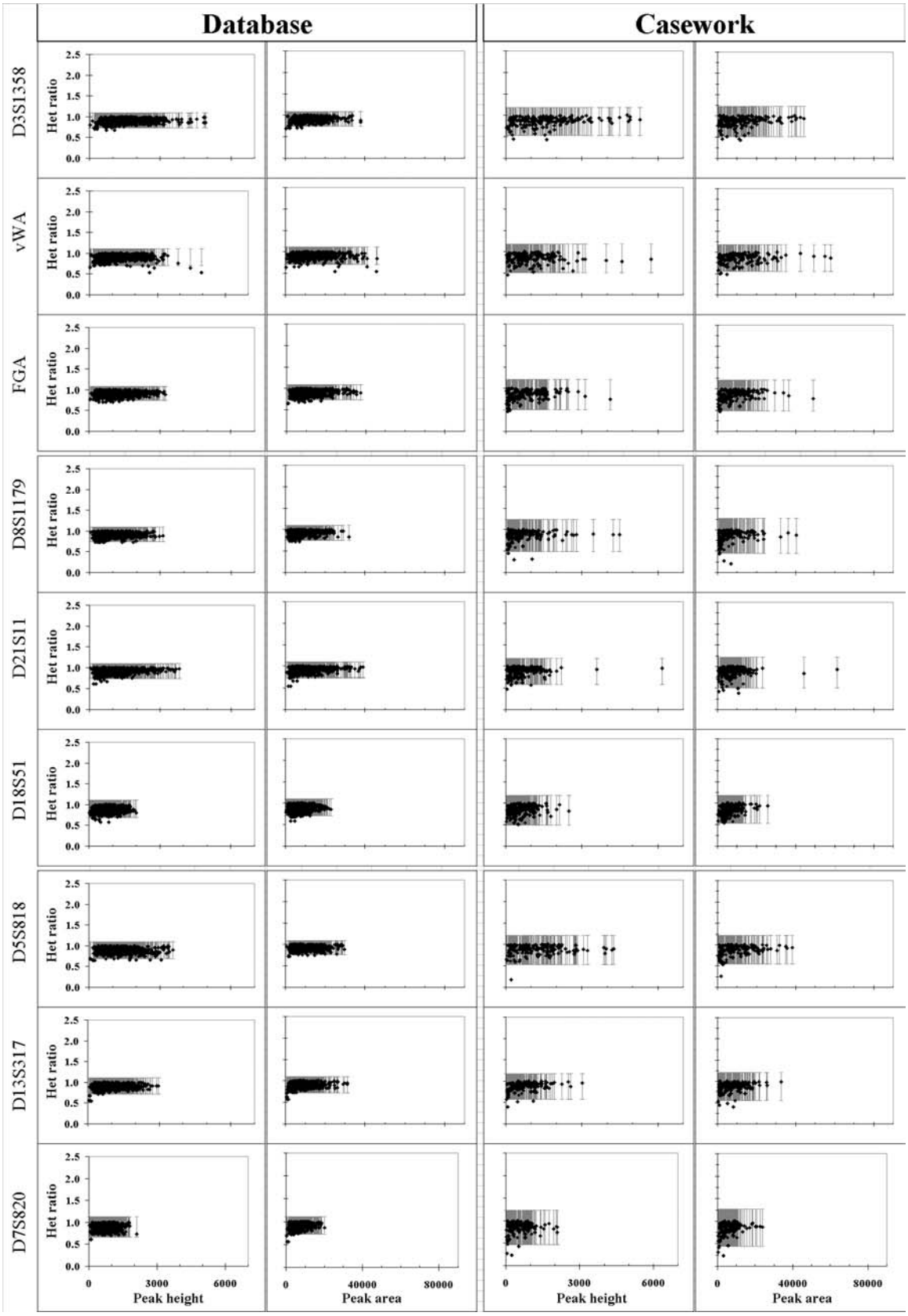

FIG. 3—*Heterozygous peak height and area ratios for database and casework samples calculated with Method #1. Method #1 calculates HR values as peak height and area of the weaker intensity allele peak over that of the stronger intensity allele peak. Stippled areas represent the range of percentage stutter covered by the average stutter value ± 3 SD for the entire locus.*

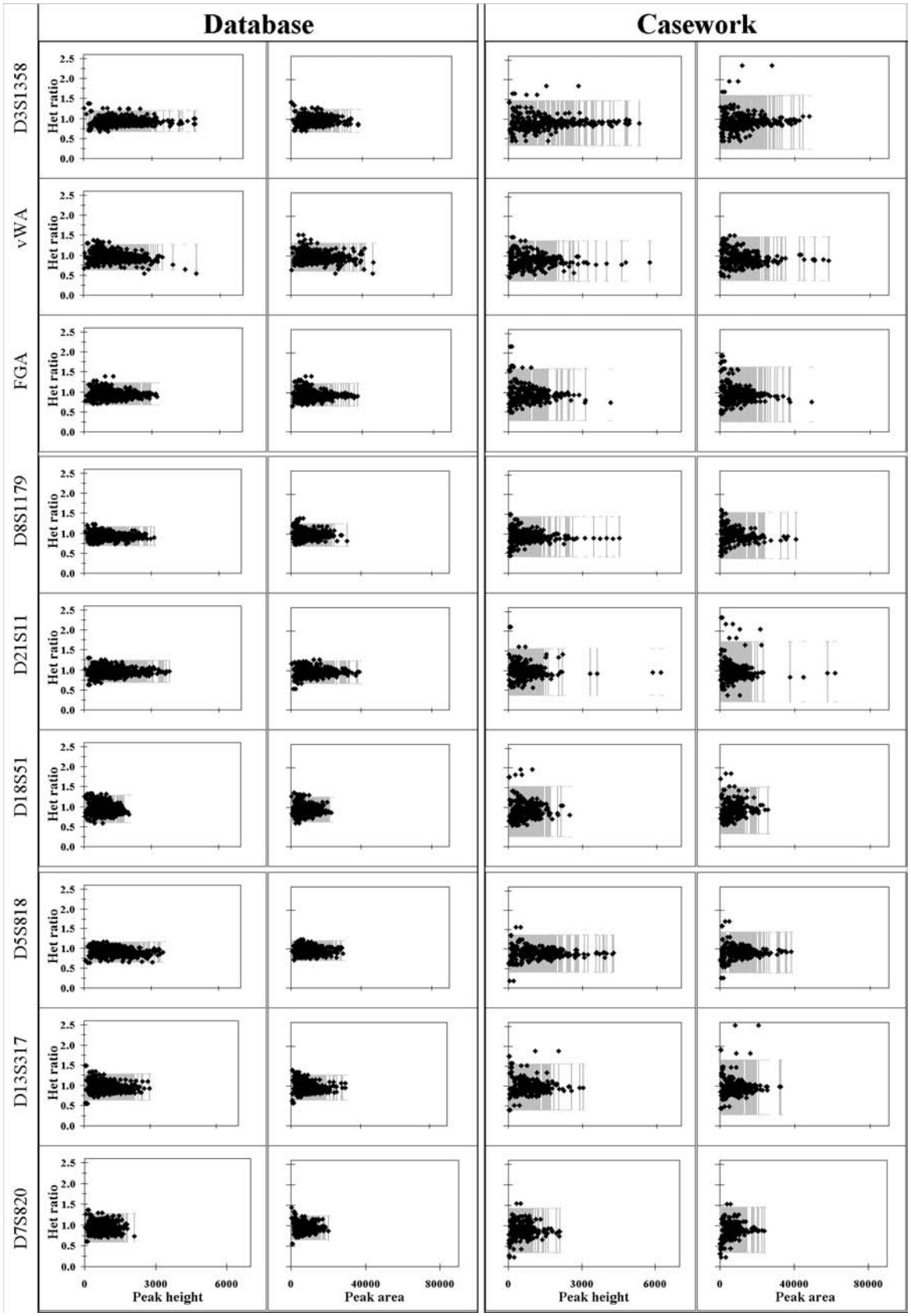

FIG. 4—*Heterozygous peak height and area ratios for database and casework samples calculated with Method #2. Method #2 calculates HR values as peak height and area of the longer allele over that of the shorter allele. Stippled areas represent the range of percentage stutter covered by the average stutter value ± 3 SD for the entire locus.*

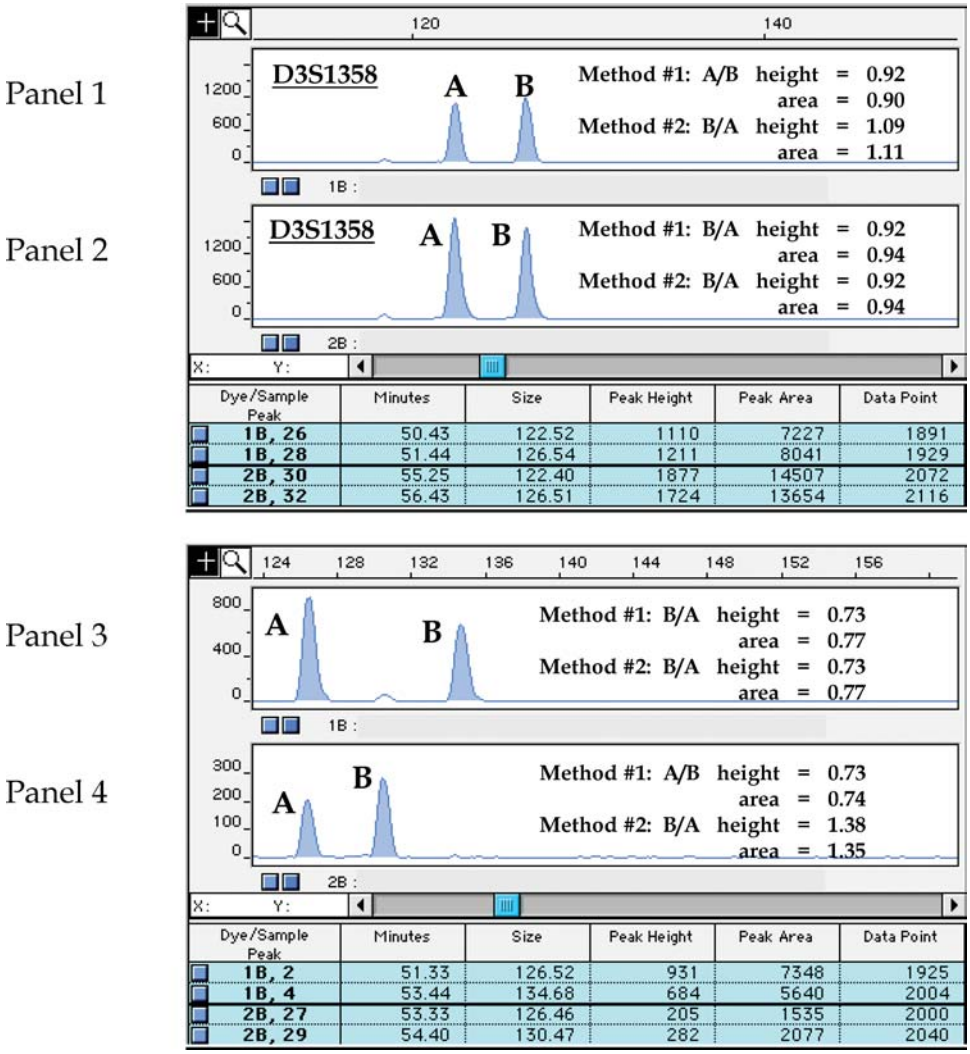

FIG. 5—*Comparison of two HR calculation methods. Samples were selected to showcase the difference in peak imbalance detection capabilities between Methods #1 and #2. Method #2 provides information on the direction of the imbalance.*

in Fig. 5. For the D3S1358 locus, panel #1 displays the smaller sized amplicon with the lower signal intensity, panel #2 displays the larger amplicon with lower signal intensity. Under the premise that the shorter amplicon is at a competitive advantage, only panel #2 should be considered as representing a near normal balance between peaks, and both calculation methods produce identical ratios. For the situation in panel #1, only Method #2 detects that the situation depicted is different from that in panel #2. This point is further emphasized with samples displayed in panels #3 and #4, where two very different situations are considered equivalent under Method #1, but clearly detected as being different by Method #2.

Figure 6 shows an example of a profile with several loci demonstrating some level of imbalance. The highlighted allelic pair features the larger allele as the highest peak intensity in the allelic pair. Even though this chromatogram does not show obvious baseline artifacts at this y-axis range, calculated ratios using peak area data significantly exceed the value of calculated ratios using peak height data. This effect can be further compounded, as shown for the sample in Fig. 7, where an incomplete A addition yielded, for the D3S1358 locus, a split peak detected for one allele only, affecting peak area based HR calculations considerably more than calculated values using peak height data. Situations where peak area was at a disadvantage in HR calculations were relatively rare in both casework and population database datasets, as evidenced by the absence of a significant difference in data point spreads between the peak area and peak height datasets.

Of the two calculation methods used to produce the HR values, Method #1 produced lowered median and SDs. This was largely expected from the theoretical performance of these methods. Figure 8 demonstrates the impact of the choice of calculation method on a series of theoretical peak ratios extending from 1:45 to the converse 45:1 for a given pair of allele peaks, A and B. The main difference between the two methods is the data point range produced by the calculations: 0 to infinity for Method #2, but a much more restricted 0 to 1 for Method #1. Figure 8 demonstrates the difference in weight given to a series of theoretical incremental changes in HR values. For both methods, HR values under 1 produce a smooth slope that gives less weight to outlying data points. However, only Method #2 produces ratios above 1 in value as under Method #1 ratios of A over B yielding a ratio over 1 in value are converted into B over A ratios and thus remain under 1 in value. Under Method #2, ratios above 1 are given more weight than once inverted as per Method #1. For example, as shown in the figure insert of Fig. 8, a 2:1 ratio under Method #2, which computes to a 100% differential, is converted under Method #1 to a 1:2 ratio, which computes to a 50% differential. It is therefore expected that median, average and standard

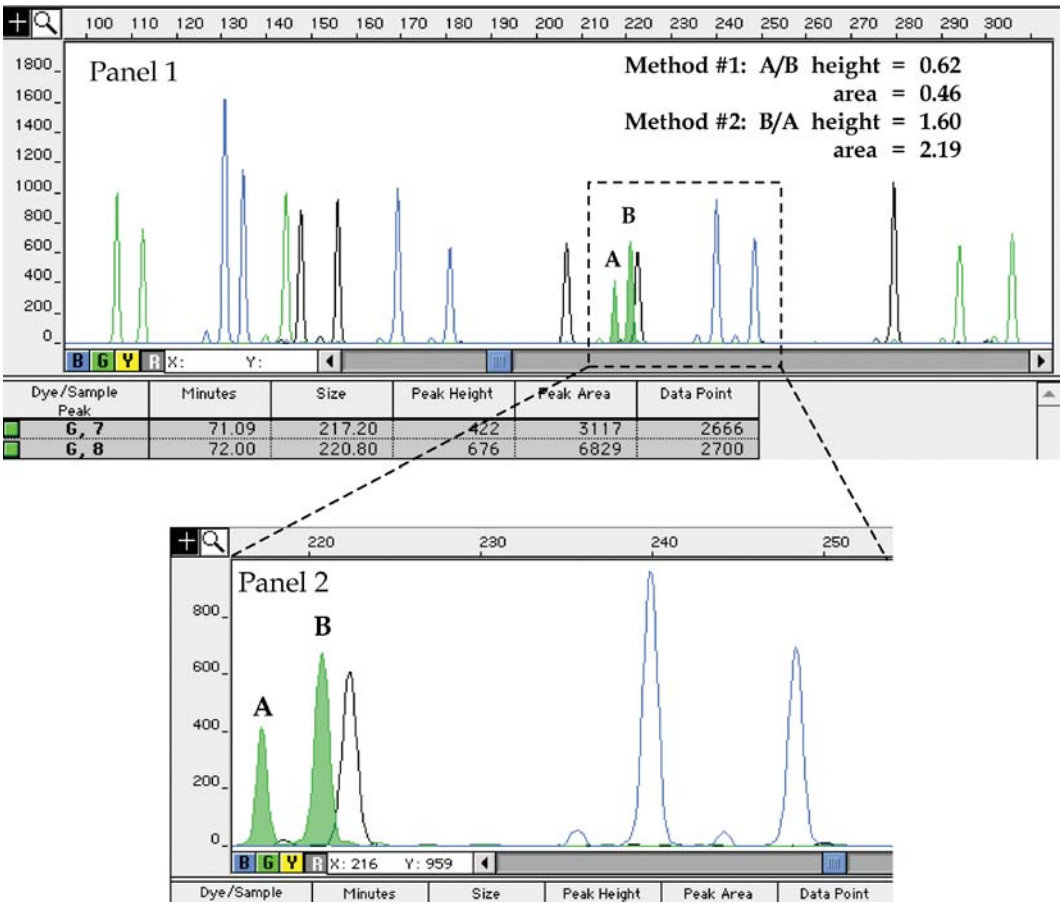

FIG. 6—*Impact of electropherogram anomalies on HR calculations. This sample demonstrates a significant discrepancy between peak height and peak area based calculations that might be due to a problem in peak area integration. Panel #2 is an enlargement of boxed area of Panel #1.*

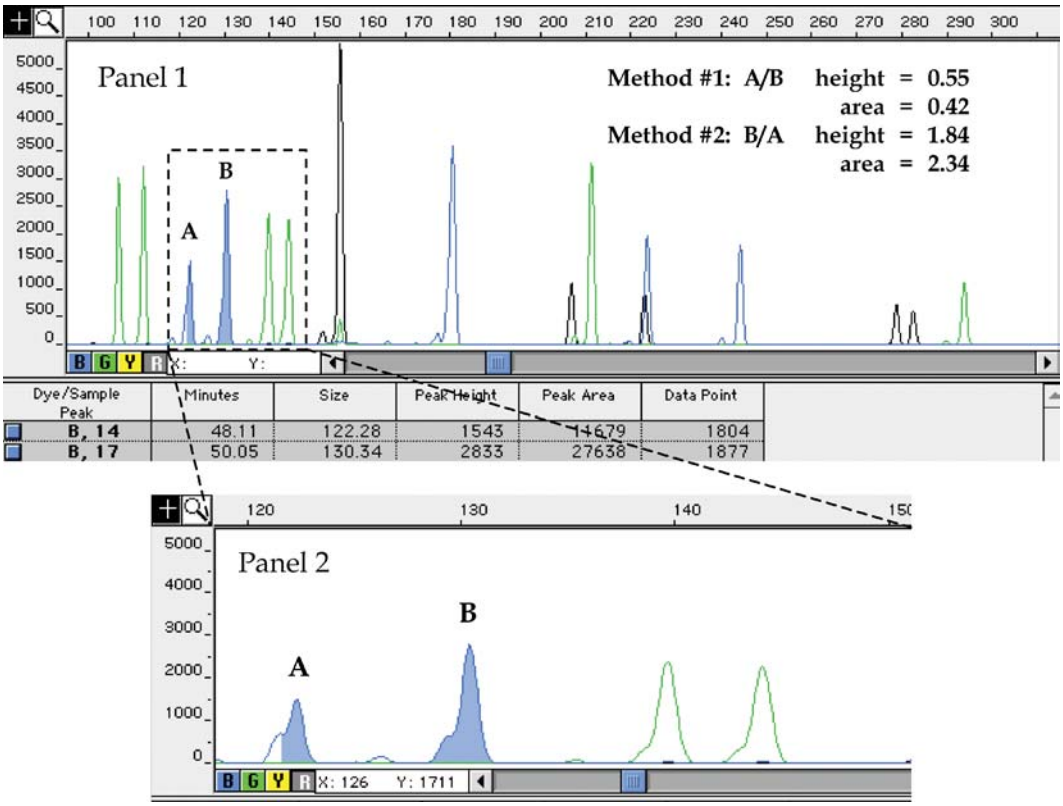

FIG. 7—*Example of the vulnerability of HR calculations using peak area. Algorithms calculating surface area under a peak can include or exclude a shoulder. This sample displays peak splitting from incomplete A addition. This shows the impact of software exclusion of one shoulder but not the other. Panel #2 is a magnification of boxed area from Panel #1.*

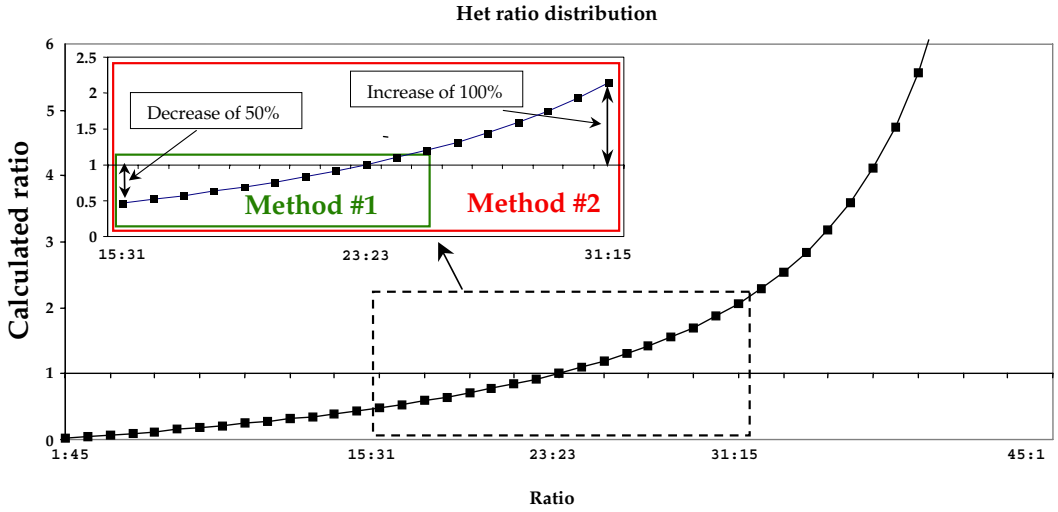

FIG. 8—*Comparison of two HR calculation methods on the theoretical peak ratios extending from 45:1 to 1:45. The arrows point to the 31:15 peak height ratio being considered under both method calculations. Under Method #1, where the ratio is calculated as lowest peak height over largest peak height, a 31:15 ratio is evaluated as a 15:31 ratio. For this 31:15 ratio, the arrows point to the 100% increase for Method #2 and 50% decrease for Method #1 in HR value, respectively, when compared to the optimal ratio of 1.*

deviations will be larger under Method #2. Method #1 provides a more even weight for each incremental change in HR values, and gives less weight to significant outliers, which is statistically preferable.

In summary, the HR values were calculated as lower signal intensity allele/higher signal intensity allele (Method #1), and as peak height and area of the longer allele over that of the shorter allele (Method #2). Method #1 was shown to provide more even weight for every ratio increment, even though information about the direction of the imbalance is lost. No significant differences were noted between peak height and peak area calculated HR values, and between loci. HR medians were approximately $93 \pm 6.5\%$ for database samples, approximately $88\% \pm 12\%$ for casework samples. The increase in standard deviation for casework samples is likely largely attributable to the challenged nature of most casework samples. At 3 SDs under the median value, the lower limit of the range under which two peaks should be considered as possibly emanating from separate profiles would be 73% and 52% for population database and casework samples, respectively. The lower limit for population database samples is in agreement with other published data (5–7,9,10). However, the lower limit for casework samples is lower, which clearly shows the importance of selecting the proper reference database for the establishment of HR threshold values for casework mixed profile interpretation. As with stutter %, these ranges should be considered as guidelines, and only once the analysis of data from all loci is completed can an assessment be made of whether or not a sample reflects the presence of a mixture.

References

1. Gill P, Sparkes R, Pinchin R, Clayton T, Whitaker J, Buckleton J. Interpreting simple STR mixtures using allele peak areas. Forensic Sci Int [PubMed] 1998;91:41–53.
2. Clayton TM, Whitaker JP, Sparkes R, Gill P. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Sci Int [PubMed] 1998;91:55–70.
3. Evett IW, Gill PD, Lambert JA. Taking account of peak areas when [PubMed] interpreting mixed DNA profiles. J Forensic Sci 1998;43:62–9.
4. Walsh PS, Fildes NJ, Reynolds R. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucl Acids Res 1996;2(14):2807–12.
5. Wallin JM, Buoncristiani MR, Lazaruk KD, Fildes N, Holt CL, Walsh PS. TWGDAM validation of the AmpFℓSTR Blue PCR amplification kit for forensic casework analysis. J Forensic Sci 1998;43(4):854–70. [PubMed]
6. Holt CL, Stauffer C, Wallin JM, Lazaruk KD, Nguyen T, Budowle B, Walsh PS. Practical applications of genotypic surveys for forensic STR testing. Forensic Sci Int 2000;112(2–3):91–109. [PubMed]
7. Lazaruk K, Wallin J, Holt C, Nguyen T, Walsh PS. Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing. Forensic Sci Int 2001;119(1):1–10. [PubMed]
8. Wallin JM, Holt CL, Lazaruk KD, Nguyen TH, Walsh PS. Constructing universal multiplex PCR systems for comparative genotyping. J Forensic Sci 2002;47(1):52–65. [PubMed]
9. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B.Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci 2001;46(3): 647–60. [PubMed]
10. Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS. TWGDAM validation of AmpFℓSTR PCR amplification kits for forensic DNA casework. J Forensic Sci 2002;47(1):66–96. [PubMed]
11. Frégeau CJ, Bowen KL, Leclair B, Trudel I, Bishop L, Fourney RM. AmpFℓSTR® Profiler Plus™ short tandem repeat DNA analysis of casework samples, mixture samples and non-human DNA samples amplified under reduced PCR volume conditions (25 μL). J Forensic Sci 2003;48:1014–34. [PubMed]
12. Royal Canadian Mounted Police, Forensic Laboratory Services Directorate. Biology Section Methods Guide. Rev. ed. Ottawa, ON, RCMP, 1998.
13. Leclair B, Fourney, RM. Automated fluorescent STR technology: simple modifications to ABD's Profiler-II protocol allow for enhanced performance. In: Proceedings of the Eighth International Symposium on Human Identification; 1997 Sept 17–20; Scottsdale (AZ). Madison (WI): Promega Corporation, 1998.
14. Leclair B, Sgueglia JB, Wojtowicz PC, Juston AC, Frégeau CJ, Fourney RM. STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. J Forensic Sci 2003;48: 1001–13. [PubMed]
15. Levinson G, Gutman GA. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol Biol Evol 1987;4(3):203–21. [PubMed]
16. Murray V, Monchawin C, England PR. The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. Nucleic Acids Res 1993;21(10):2395–8. [PubMed]
17. Hauge XY, Litt M. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. Hum Molec Genet 1993;2(4):411–5. [PubMed]

18. Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. Nucleic Acids Res 1992;20(2):211–5. [PubMed]
19. Blackmore VL, Luebke MH, Laird JM, Newall PJ. Preferential amplification and stutter observed in population database samples using the AmpFℓSTR Profiler multiplex system. Can Soc Forens Sci J 2000;33(1):23–32.
20. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 1988;239:487–91. [PubMed]
21. Innis MA, Gelfand DH. Optimization of PCRs. In: PCR Protocols: A Guide to Methods and Applications. Academic Press, 1990;3–12.
22. Ruano G, Brash DE, Kidd KK. PCR: The first few cycles. Amplifications 1991;7:1–4.
23. Ruano G, Kidd KK. Modeling of heteroduplex formation during PCR from mixtures of DNA templates. PCR Methods and Applications 1992;2:112–6. [PubMed]
24. Bell DA, DeMarini DM. Excessive cycling converts PCR products to random-length higher molecular weight fragments. Nucleic Acids Res 1991;19(18):5079. [PubMed]

Additional information and reprint requests:
Ron M. Fourney, Ph.D.
National DNA Data Bank
Royal Canadian Mounted Police
1200 Vanier Parkway
Ottawa, Ontario K1G 3M8
Canada