# STR ANALYSIS BY CAPILLARY ELECTROPHORESIS: DEVELOPMENT OF INTERPRETATION GUIDELINES FOR THE PROFILER PLUS™ AND COFILER™ SYSTEMS FOR USE IN FORENSIC SCIENCE

**Deborah Hobson, Jill Smerick, and Jenifer Smith**

*DNAU I, FBI Laboratory, Washington, D.C.*

This paper discusses aspects of short tandem repeat (STR) analysis that are considered in profile interpretation. These include thresholds, peak height ratios, stutter and incomplete non-template nucleotide addition. Together, these features are used in the interpretation of STR profiles.

**Thresholds and Peak Height Ratios**

Thresholds can be divided into two categories: analytical and match. An analytical threshold is the level that collected data is quantitatively evaluated. A match threshold is the level that analyzed data is used for match purposes. The analytical threshold can be chosen as the detected signal above baseline noise. Alternatively, an analytical threshold can be the same value chosen as the match threshold. A match threshold should be chosen based on studies that evaluate stochastic amplification. As such, a match threshold is the signal at or above which stochastic amplification is minimal, and will not result in allelic dropout. Allelic dropout occurs when one allele of a pair exhibits signal, whereas the other peak falls below detection. It can be considered "the line in the sand" for match purposes. Data falling below this threshold may not be suitable for match purposes; however, it can be used to exclude. Data at or above the match threshold, however, is of sufficient quality to determine a match. In choosing a match threshold, it should be remembered that forensic scientists must be both objective and conservative. Matches must be based on high quality data.

The past PCR systems of AmpliType® DQA1 and Polymarker were developed with the consideration of stochastic amplification (1, 2). Hence, these systems were evaluated prior to determining a match threshold. In reviewing FBI quality control (QC) data that spanned several years, the quantity of DNA present in which a "S" or "C" dot was present (i.e., equivalent to a match threshold) was approximately 250 pg DNA. This is approximately 80 copies of DNA, a sufficient number of templates to limit stochastic amplification and hence, avoid allelic dropout. This data was considered during the match threshold development phase.

A study was performed to: 1) determine a match threshold which would safely avoid stochastic effects that could lead to allelic dropout, and 2) develop peak height ratio guidelines to assist in mixture analysis (3). DNA extracts from 25 individuals were serially diluted from 1 ng to 63 pg DNA and amplified with the AmpF*l*STR® Profiler Plus®™ and Cofiler™ kits (4, 5). The amplicons were analyzed on five ABI PRISM™ 310 Genetic Analyzers (6). The Peak Amplitude Threshold (PAT) in Gene® Analysis was set to capture all data (7). The allelic designations and peak heights were captured in GenoTyper®, and this data was imported into Excel where peak height ratios were determined for heterozygotes at each locus (8). The peak height ratio was defined as:

(A/B) X 100%

where A is the peak height in relative fluorescent units (RFU) of the shortest peak and B is the peak height (RFU) of the tallest peak for a pair of heterozygote alleles at a locus. The data from this study are presented in Figures 1 through 3, Figure 5 and Table 1 below.

Figure 1 demonstrates that although most of the peak height ratios are greater than 70%, there is an increase in peak ratio variation with a corresponding decrease in peak height. As figure 2 shows, peak height ratio is a function of template DNA. This is expected given the nature of the PCR reaction. Figure

2 also demonstrates the necessity of using serially diluted DNA samples to ascertain peak height ratios. The sole use of 1 ng samples, for example, would have produced erroneous peak ratios.

The data from this study in which one allele's peak height is greater than 50 RFU and one allele's peak height is less than 50 RFU was termed "straddle data". Table 1 shows the average peak height ratio of straddle data is 50 RFU or slightly greater, with a large standard deviation. Thus, the possibility for allelic dropout increases at low peak heights.

Table 1. Peak height ratios of samples in which one allele falls below 50 RFU

| 310 Genetic Analyzer | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Average | 50 | 54 | 56 | 51 | 57 |
| Standard Deviation | 15 | 20 | 21 | 13 | 20 |

Indeed, there were two occurrences of allelic dropout in this study. In one of these occurrences, both alleles of a heterozygote at the D18S51 locus were visible on one 310 Genetic Analyzer with peak heights of 98 and 28 RFU. When this sample was analyzed on a 310 Genetic Analyzer with slightly lower detection capabilities, the peak height of the tallest allele was detected as 65 RFU while the shortest peak was not detected above noise.

Another example of straddle data is presented in Figure 3. This figure depicts results at the D3S1358 locus from 1 ng (upper panel) and 63 pg (lower panel) template quantities exhibiting 81% and 30% peak height ratios, respectively. Although, panel B does not demonstrate allele dropout, the peak detected as 46 RFU would not be recognized by GenoTyper® unless the GeneScan® PAT was adjusted.

The FBI's analytical and match thresholds are 50 and 200 RFU (9). Data that falls between 50 RFU and 200 RFU are used for exclusionary purposes, and to determine the presence of a mixture. Table 2 shows the peak height ratio guidelines employed at the FBI.

Table 2: FBI Peak Height Ratio Guidelines for Profiler Plus and Cofiler Loci

| | 200 to 300 RFU | 300 to 1000 RFU | > 100 RFU |
|---|---|---|---|
| Profiler Plus | 55 to 60% | 60 to 65% | 65 to 70% |
| Cofiler | 60% | 60 to 65% | 70 to 75% |

Analyzing data between 50 and 200 RFU shows the presence of low level spikes, bleed-through, tails, dissociated dye, low level incomplete non-template nucleotide addition, most all stutter peaks, and the presence of a low level second contributor(s). However, the advantages of capturing this data are that it can be characterized and presented in case notes. In addition, low level contamination would be clearly visible and corrective action could be initiated. Most importantly, this data can be used to exclude a contributor, determine the presence of a mixture, or to indicate a potential inclusion that requires an increase in DNA template (for amplification) to increase the signal above the match threshold.

Stutter

Stutter is expected and well defined (10). With respect to single source samples, stutter is not a significant occurrence. However, in mixture interpretation, percent stutter guidelines can be used to determine if an allele may exist at the N-4 position of another allele (N).

Our laboratory evaluated percent stutter of Profiler Plus loci for samples amplified with the Perkin Elmer GeneAmp 9700 using 1 ng DNA samples from 25 individuals. This data was compared to that produced by the FBI's Forensic Science Research and Training Center (FSRTC) (11). The FSRTC's stutter evaluation was perfomed on the Perkin Elmer GeneAmp 9600s (ref). See Figure 4. As demonstrated, the data is not significantly different. The data from FSRTC represents a large data set, hence, slightly greater variation is expected.

Additionally, as part of the study discussed above (see Thresholds and Peak Height Ratios) stutter data was evaluated. As Figure 5 shows, stutter peaks greater than expected maximums can occur. In this

data set, enhanced stutter was rare, however it can accompany samples with low peak height indicating that it may occur randomly with low copy number samples.

The DNAU I utilizes the FSRTC data generated for stutter as a guide for the maximum expected stutter. As Figure 4 depicts, these guidelines are locus specific. However, because enhanced stutter may be associated with low peak height, these values are used solely as guidelines.

Non-Template Nucleotide Addition

Non-template nucleotide addition is an expected PCR phenomenon which can be regulated by primer design and extension times (12). Incomplete non-template nucleotide addition ("minus-A") results in a peak one base smaller (N-1) than the primary allele peak (N). The Profiler Plus and Cofiler typing kits are designed to achieve near-complete adenylation of PCR product. However, over-amplification due to excess DNA template can result in disproportionate minus-A, resulting in potential interpretation difficulties. Human DNA quantification and the subsequent control of DNA template used for amplification can resolve this potential limitation.

A study with DNA ranging from 0.5 to 4 ng was conducted to determine the percent minus-A obtained and the difference in size from the primary peak. At these quantities of template DNA, the amount of incomplete non-template nucleotide addition was minimal (a range of 4 to 21%) and appears as shoulder peaks adjacent to the primary peaks (See Figure 6). The range of size difference was 0.9 to 1.06 bases. It was observed that minus A occurred primarily at vWA and D3S1358. Furthermore, minus-A occurred in both alleles of a heterozygote individual in only 3 out of 25 occurrences.

However, when an excess of DNA is amplified (i.e., greater than 4 ng), Taq polymerase is unable to efficiently adenylate PCR product. This results in a significant percent of incomplete non-template nucleotide addition (see Figure 7). Note that minus-A occurs in both pairs of alleles at both loci. Amplifying these samples with less DNA will rectify this situation.

In single source samples incomplete non-template nucleotide addition is noted by demonstrating the expected criteria (locus, percent, difference in size). These same criteria are used in mixture interpretation to determine if a peak is incomplete non-template nucleotide addition, or an allele. When a significant peak is present in the N-1 position, it is important to determine: 1) if the data is off-scale (greater than 8191 RFU in the raw data), 2) which locus is at issue, 3) if there are unaccounted alleles, and 4) if an allele 1 base smaller than N is possible (i.e., TH01 9.3 and 10 alleles).

Acknowledgements

The authors thank Bruce Budowle, Tamyra Moretti, Alice Brown, Anne Baumstark, Debra Defenbaugh and Kathleen Keys for their assistance with STR implementation.

References

AmpliType User Guide (September 1990). Cetus Corportation, Emeryville, California.

Budowle, B., Lindsey, J.A., DeCou, J.A., Koons, B.W., Giusti, A.M., and Comey, C.T., "Validation and Population Studies of the Loci LDLR, GYPA, HBGG, D7S8 and Gc (PM Loci), and HLA-DQα Using a Multiplex Amplification and Typing Procedure,"J. For. Sci., 40(1):45-54, 1995.

Hobson, D.L., Duer, D., Smerick, J.B., Smith, J.A., Reeder, D., and Budowle, B., "Establishing Peak Height Ratio Guidelines and An Interpretation Threshold for the AmpF/STR® Profiler Plus™ and Cofiler™ DNA Typing System on the ABI Prism™ 310 Genetic Analyzer,"In preparation.

AmpF/STR® Profiler Plus™ PCR Amplification Kit User's Manual, 1998.

AmpF/STR® COfiler™ PCR Amplification Kit User Bulletin, 1998.

ABI Prism™ 310 Genetic Analyzer User's Manual, 1998.

ABI Prism™ GeneScan® Analysis Software 2.1 User's Manual, September 1996.

ABI Prism™ Genotyper® 2.0 User's Manual, 1996.

Short Tandem Repeat Analysis Protocol, FBI Laboratory, January 1999.

Walsh, P.S., Fildes, N.J., and Reynolds, R., "Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA," Nucleic Acid Res., 1996, **24**, 2807-2812.

Moretti, T.R., Baumstark, A.M., Brown, A.L., Defenbaugh D.A., Keys, K.M. and Budowle, B., "Validation of STR Typing by Capillary Electrophoresis," submitted to the Journal of Forensic Sciences.

Kimpton, C., Gill, P., Walton, A., Urquhart, A., Milican, E., and Adams, M., "Automated DNA Profiling Employing Multiplex Amplification of Short Tandem Repeat Loci," PCR Methods and Applications, 1993, **3**:13-22.

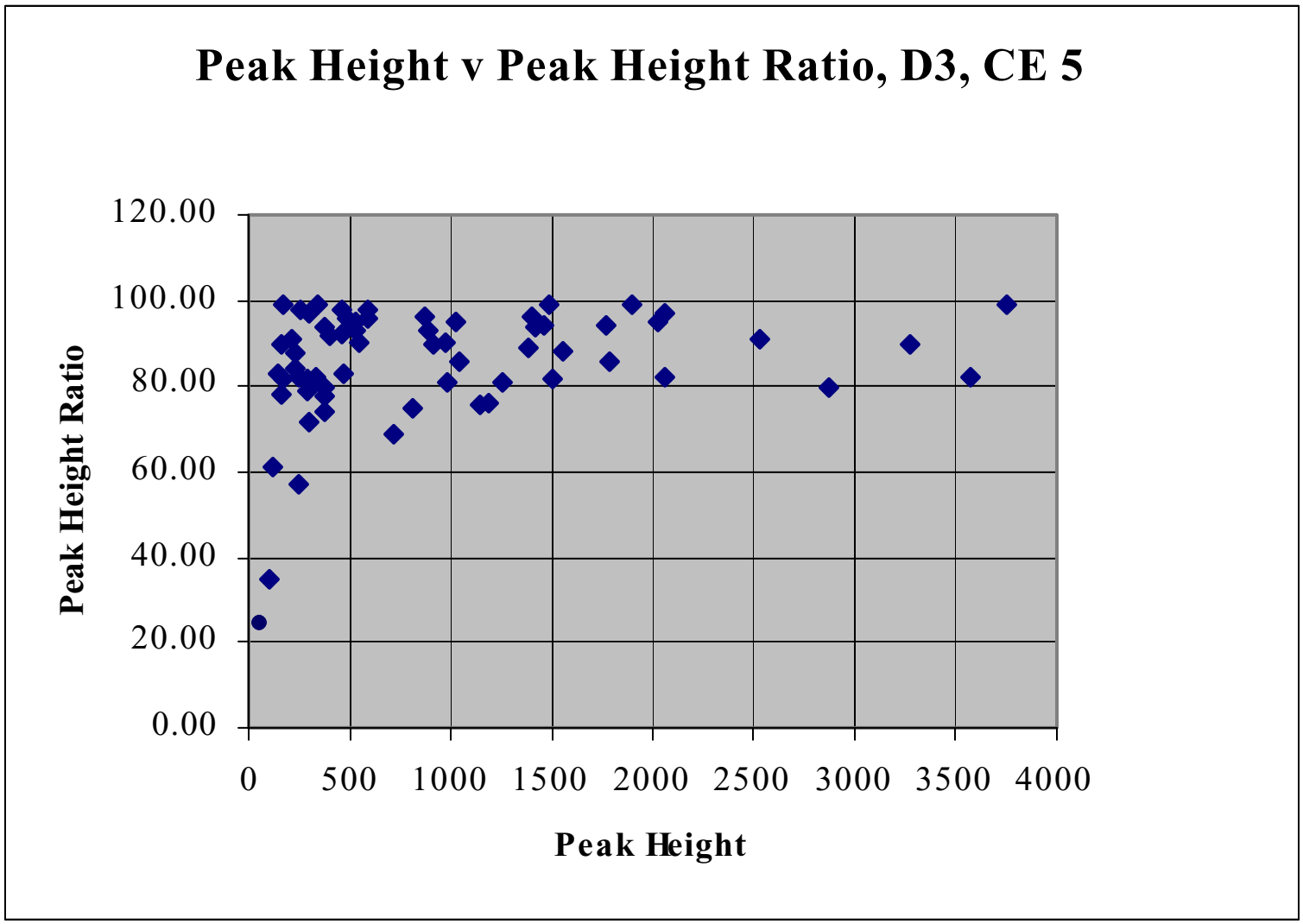



Figure 1: Peak height ratios ofheterozygotepairs at the Profiler Plus D3S1358 locus as a function of peak height.

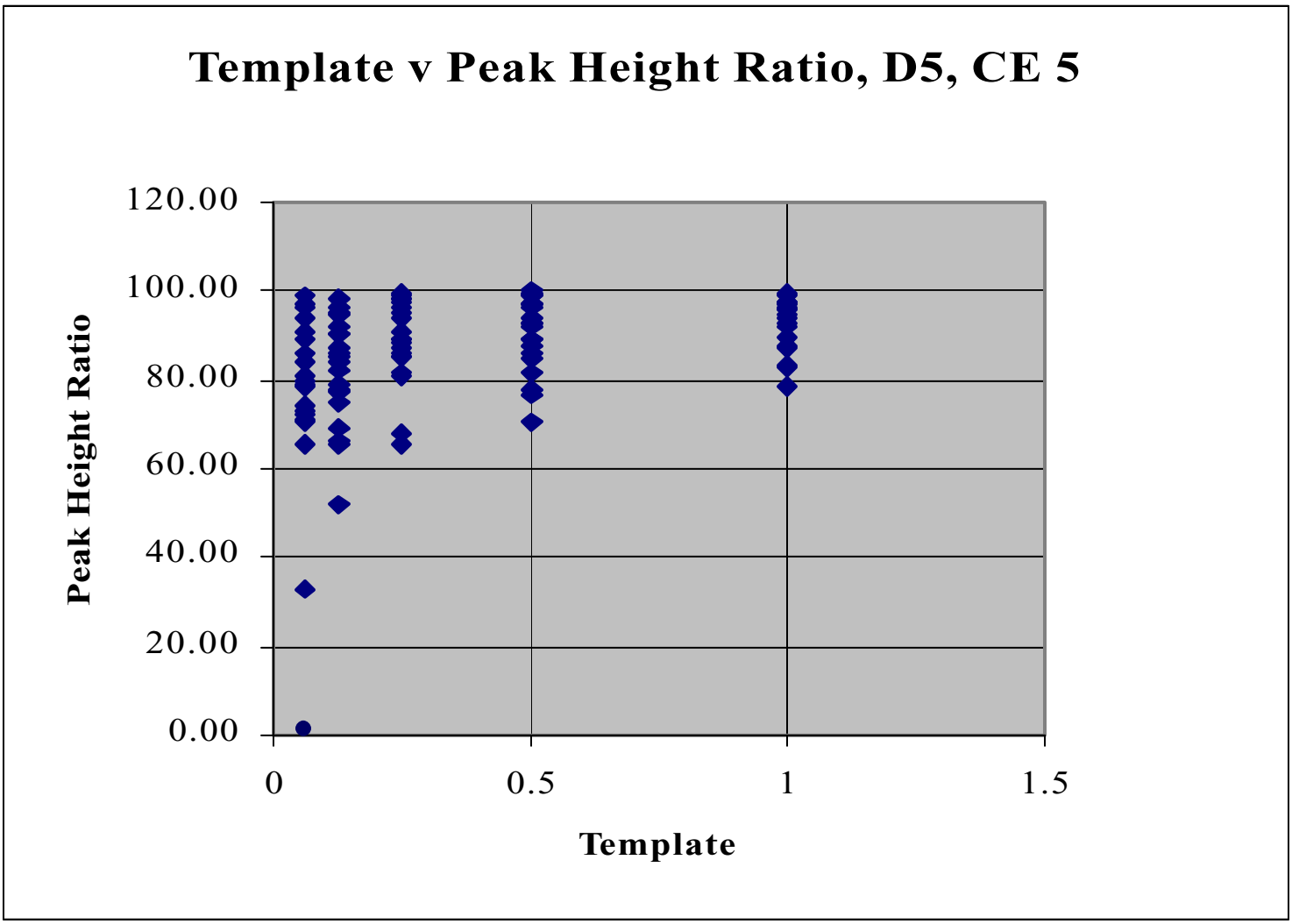



Figure 2: Peak height ratios ofheterozygotepairs at the Profiler Plus D5S818 locus as a function of template DNA (63 pg, 125 pg, 250 pg, 0.5 ng and 1 ng).

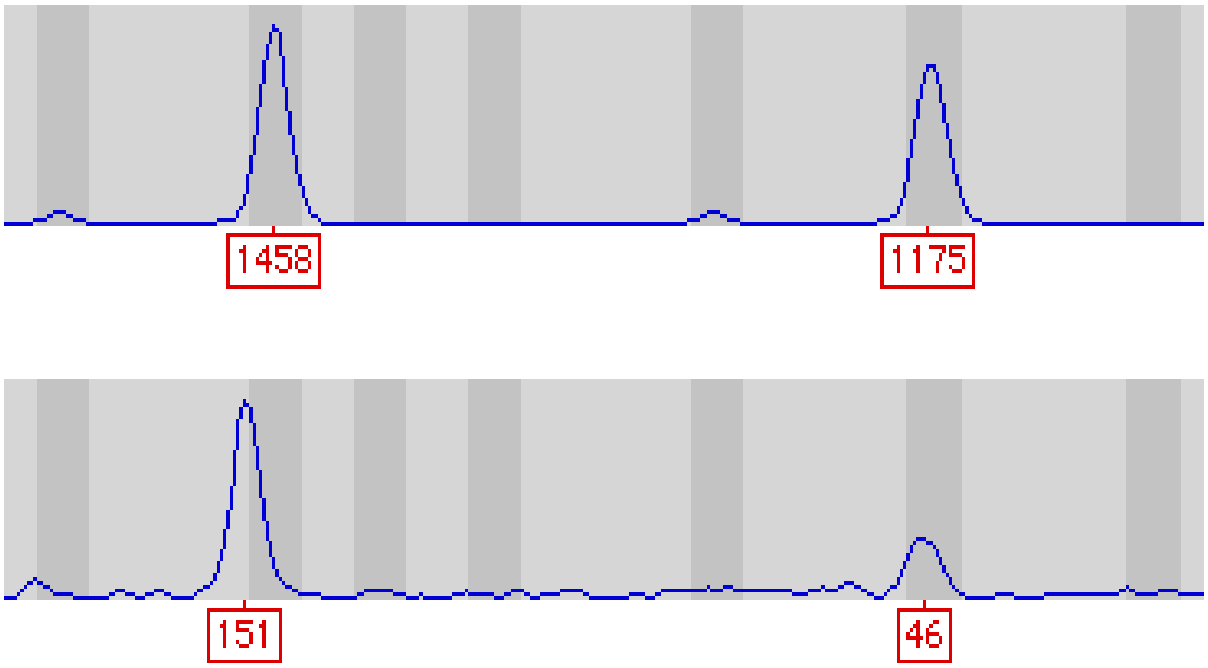



Figure 3: Peak height data at the D3S1358 locus from 1ng (panel A) and 63 pg (panel B) exhibiting 81% and 30% peak height ratios, respectively. Note that one of the alleles of the 63pg template falls below 50 RFU, which would not be labeled by GenoTyper.

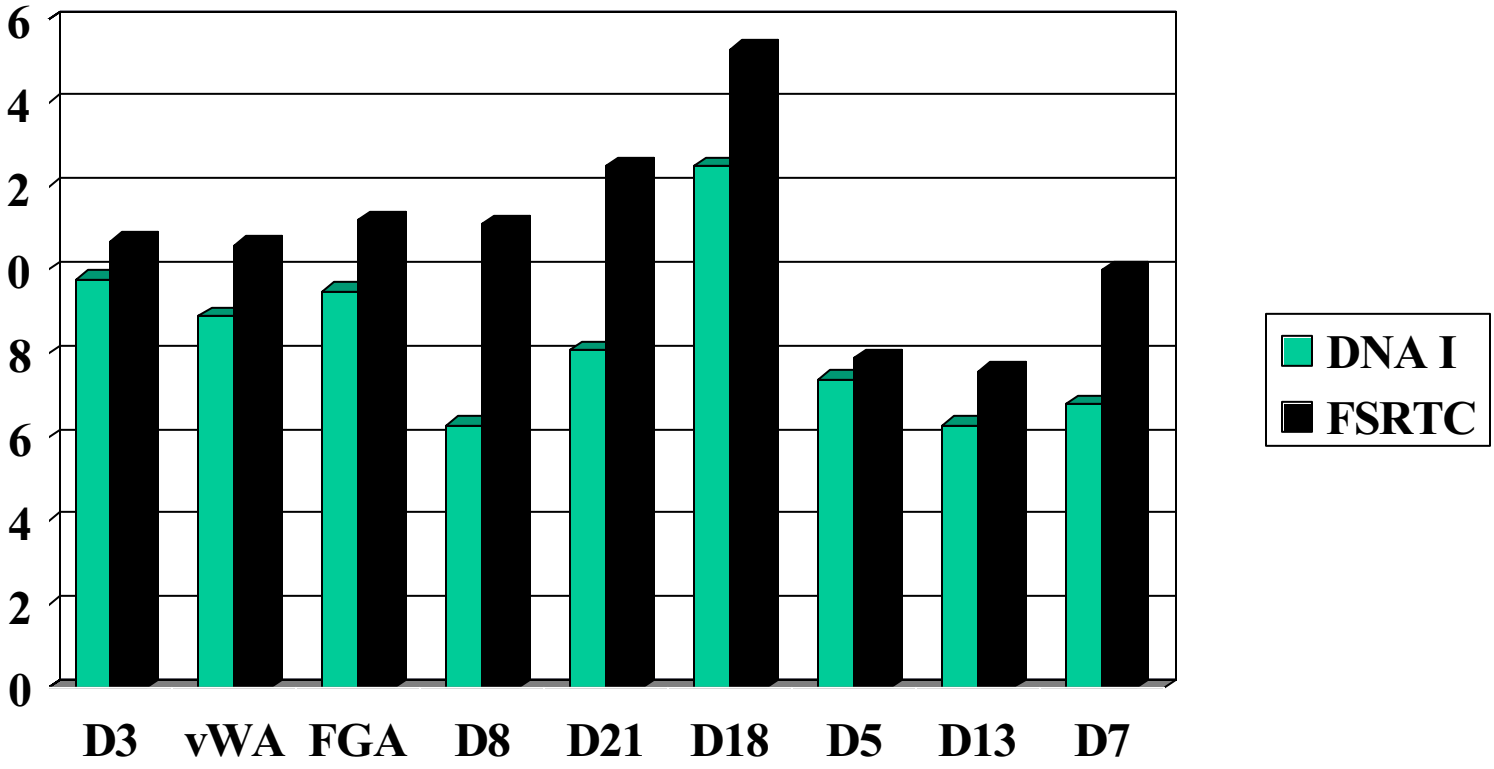
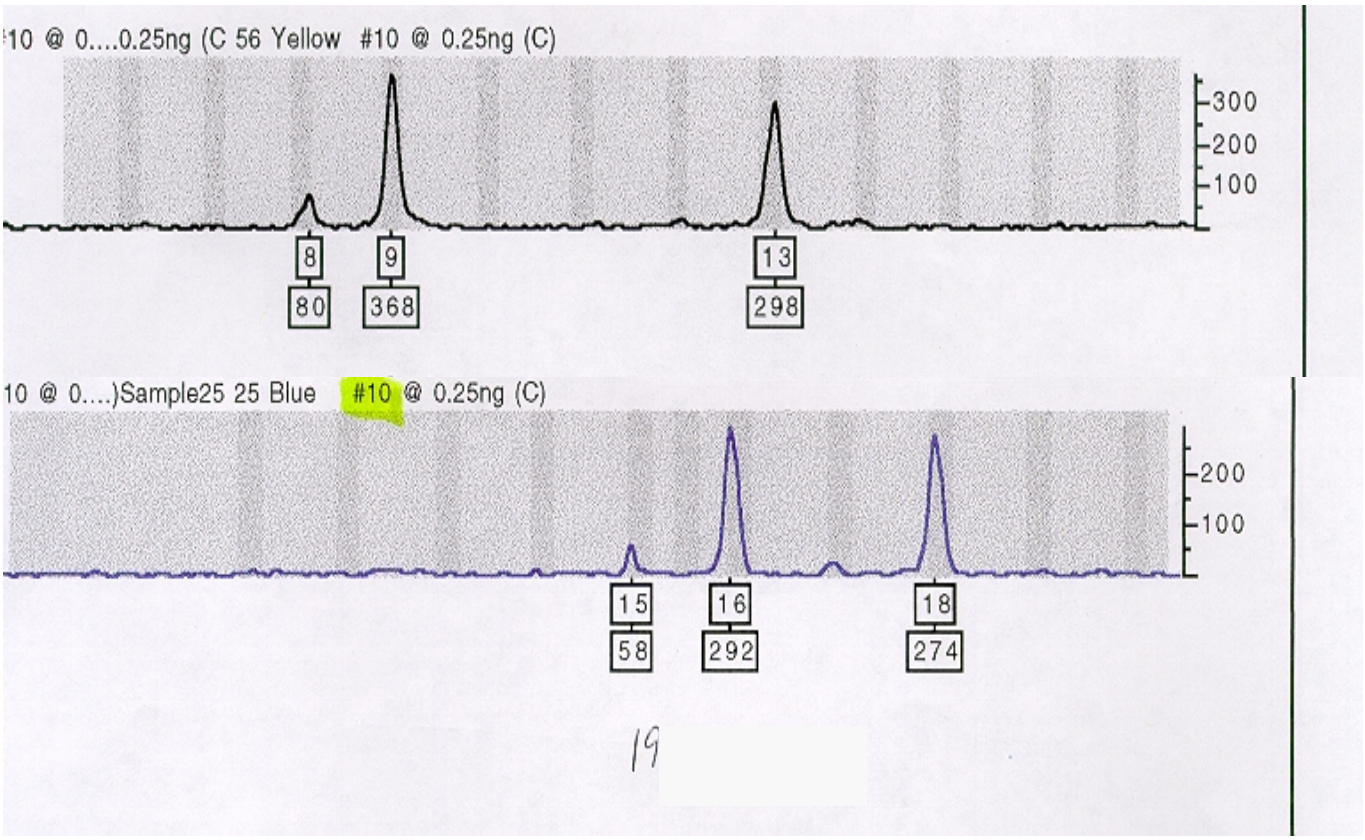



Figure 4: Stutter data for Profiler Plus loci. Each bar represents the average plus three standard deviations.




A

B

Figure 5: Stutter data that exceeds the expected maximum value (Figure 4). Panel A displays 22% stutter of the "9" allele at the D5S818 locus. Panel B displays 20% stutter of the "16" allele at the D3S1358 locus. Note the low peak height of these alleles.

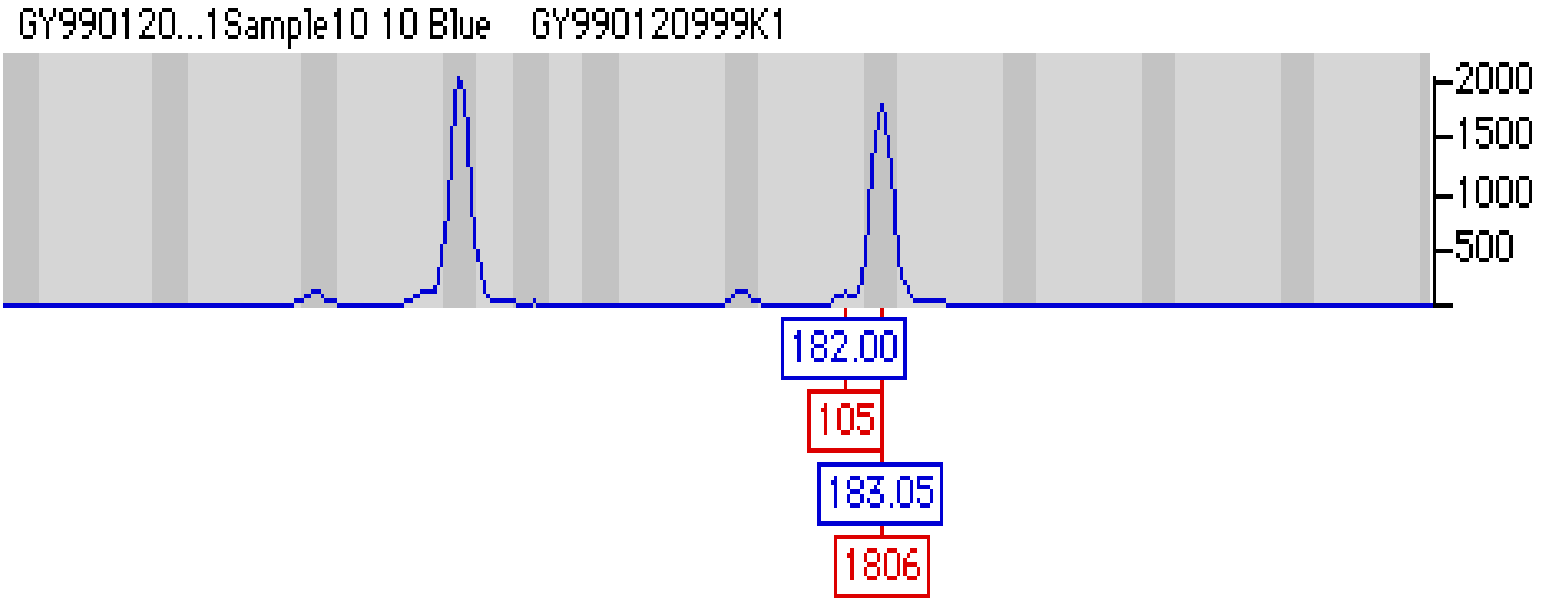



Figure 6: Incomplete non-template nucleotide addition at the vWA locus. The N-1 peak is 6% of the N peak, and is sized 0.95 bases smaller. Note that this data is "on scale."

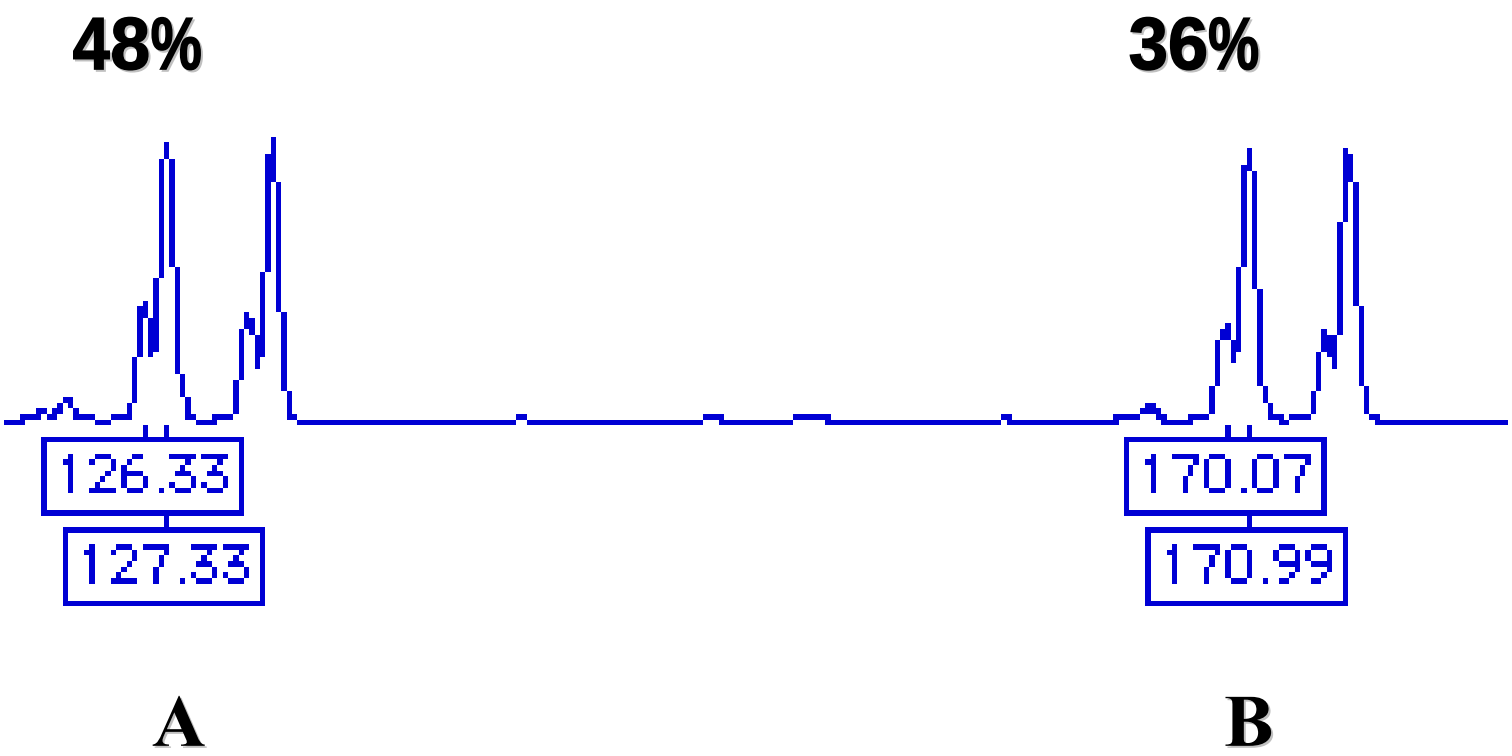



Figure 7: Incomplete non-template nucleotide addition. The N-1 peak is 48% at the D3S1358 locus (A), and 36% at thevWA locus (B).