The Evaluation and Implementation of GeneMarker®HID (Softgenetics), PowerPlex® Fusion 6C (Promega), and STRmix™ (ESR) for Forensic Casework.

John E. Schienman, Ph.D., Angela Przech, Ph.D., Michael Morganti, Steven Bryant, Daniel Renstrom, Melanie Russell, Jillian Echard, Chia-Hung Hsiao, Ph.D., Cheryl Carreiro, Dahong Sun, M.D., Ph.D., Guy Vallaro, Ph.D., Carll Ladd, Ph.D.

State of Connecticut Department of Emergency Services and Public Protection, Division of Scientific Services 278 Colony Street, Meriden CT 06451

ABSTRACT

Adding seven loci to the CODIS core effective January 1, 2017 requires forensic science laboratories to validate and implement one of the newer megaplex STR kits. However, the additional loci may increase the time required by the analyst to analyze results and document edits. Most of the new megaplex STR kits utilize 6-dyes to accommodate the extra loci, which then necessitates forensic labs to validate new DNA analysis software to accommodate the additional dye. Finally, the forensic community is transitioning to probabilistic genotyping software to make full use of the peak heights, especially regarding DNA mixtures. Our laboratory has chosen Promega's Fusion 6C STR amplification system, Softgenetics' GeneMarker HID analysis software, and ESR's STRmix probabilistic genotyping software to address the new requirements and operational considerations.

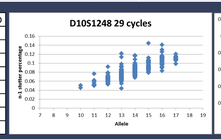
Given that n-1 repeat stutter percentage increases with allele size, the use of a single stutter threshold for a locus is not optimal. Since previous analysis software options can only have a single locus stutter threshold, the filter would either under or overcompensate at any given allele depending on the value used (locus minimum+3SD, average+3SD, or maximum) thus requiring manual edits or producing a less accurate mixture DNA profile. Another common practice is to manually add a label (e.g., an asterisk *) to the electropherogram for subthreshold peaks in a lab defined range such as above the analytical threshold, but below the stochastic threshold. Our laboratory contracted Softgenetics to modify the GeneMarker HID software to incorporate allele-specific stutter and subthreshold peak labeling options. These modifications minimize the number of manual edits required, saving analysis time and produce more accurate mixture DNA profiles. In allele-specific stutter mode, GeneMarker HID software can successfully analyze n-1, n+1, n-0.5, and n+0.5 repeat stutter positions including when stacked. Single source samples analyzed with allele-specific stutter thresholds in comparison to locus-specific had less manual edits. Peaks in overlapped stutter positions were correctly evaluated. GeneMarker HID is a cost-effective option relative to some others, significantly reducing costs for forensic laboratories with a large number of analysts. To illustrate the full DNA analysis workflow, a mock three person female-male-male mixture was prepared at a ratio of 10:4:1 with and without degradation, amplified with Fusion 6C and analyzed by GeneMarker HID. The DNA profile results were then exported to STRmix. Likelihood ratio results are presented for true contributors and non-contributors.

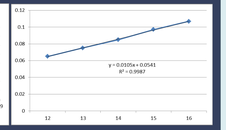
Method for Capturing Allele-Specific Stutter

1. Amplified 367 single source samples (extracted with DNA IQ™ system - Promega) with Promega's PowerPlex® Fusion 6C at 29 cycles.
2. Analyzed data with 20 RFU threshold and no stutter filters.
3. Generated an average of 366 N-1 repeat stutter data points per locus that ranged from 140 (Penta D) to 532 (D1S1656).
4. Calculated allele-specific averages and standard deviations. Plotted averages to visualize data patterns and determine slope for average stutter % change per allele (slope based on alleles with many data points).
5. Using the slope, extrapolate from average + 3SD for allele with the most data points (minus outliers).

The D10S1248 Locus has a Simple N-1 Stutter Pattern

Allele	n-1 Data Pts	Avg	St Dev	Avg+3SD
10	2	0.0480	0.0037	0.0593
11	9	0.0571	0.0083	0.0821
12	42	0.0650	0.0074	0.0871
13	139	0.0751	0.0092	0.1026
14	127	0.0848	0.0079	0.1087
15	51	0.0970	0.0107	0.1290
16	32	0.1068	0.0125	0.1442
17	13	0.1103	0.0054	0.1264



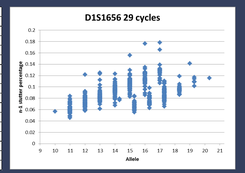


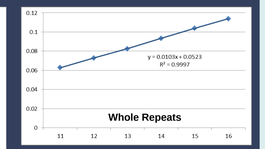
Extrapolated from average+3SD for 13 allele at 10.26% using slope of 1.05% (then rounded to nearest 0.1), e.g. 12 allele n-1 stutter % = 9.2, 14 allele n-1 stutter % = 11.3.
Stutter % for n-1 repeats held constant for alleles ≤ 10 at 7.1%.

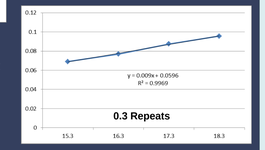
The D1S1656 Locus has a Bimodal N-1 Stutter Pattern Due to LUS* of Repeat Types

* Longest Uninterrupted Stretch, stutter % correlates with LUS

Allele	n-1 Data Pts	Avg	St Dev	Avg+3SD
10	1	0.0568		
11	49	0.0628	0.0080	0.0867
12	67	0.0735	0.0101	0.1038
13	48	0.0848	0.0119	0.1206
14	67	0.0938	0.0098	0.1231
14.3	2	0.0781	0.0015	0.0826
15	77	0.1049	0.0104	0.1362
15.3	24	0.0690	0.0086	0.0947
16	49	0.1151	0.0138	0.1566
16.3	30	0.0769	0.0075	0.0995
17	21	0.1283	0.0167	0.1783
17.3	65	0.0874	0.0097	0.1166
18.3	25	0.0956	0.0067	0.1157
19	1	0.1411		
19.3	5	0.1102	0.0058	0.1275
20.3	1	0.1155		







Whole Repeats (also N.1, N.2)

Extrapolated from average+3SD for 15 allele (minus 2 outliers) at 12.85% using a slope of 1.03% (then rounded).

N.1 and N.2 repeats use stutter % for N repeat,
e.g. stutter % for the 12, 12.1, 12.2 alleles = 9.8%
Stutter % for n-1 held constant for alleles ≤ 11.2 at 8.7%.

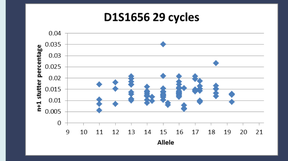
N.3 Repeats

Extrapolated from average+3SD for 17.3 allele at 11.66%, using slope of 0.9% (then rounded).
Stutter % for n-1 held constant for alleles ≤ 15.3 at 9.9%.

N+1 & N-0.5 Repeat Stutter Not Significicantly Affected by Allele Size

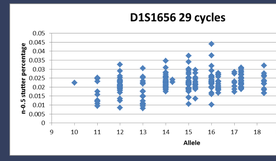
(exception N+1 Stutter for D22SS1045, a tri-nucleotide repeat)

D1S1656 N+1 Repeat Stutter



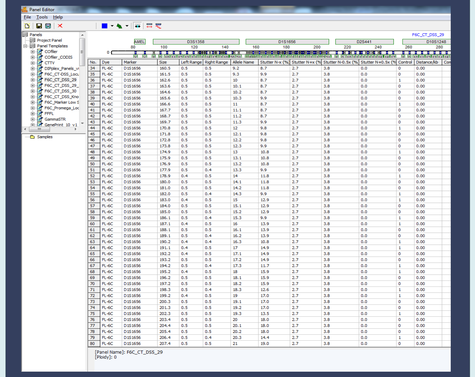
# Data Points	Overall Avg.	SD	Avg+3SD
89	0.0137	0.0043	0.0267

D1S1656 N-0.5 Repeat Stutter



# Data Points	Overall Avg.	SD	Avg+3SD
238	0.0222	0.0053	0.0381

GeneMarker® HID Panel Editor View of D1S1656 Allele-Specific Stutter (AS-ST)



Comparing the Number of Manual Stutter Edits Required for Single Source Samples Using AS-ST vs Locus-Specific Stutter (LS-ST) Thresholds

Single Source Samples (45 samples, 9 genotypes)

Number of Manual Stutter Edits

	n-1	n-1/n+1	Total	Samples w/Edits	Edits/Sample (45)
CT Allele-Specific	6	0	6	4	0.13
CT Locus Minimum	314	94	408	44	9.07
Promega Avg.+3SD	10	9	19	14	0.42

Allele-Specific Stutter Thresholds Minimize Manual Edits.

Comparison of Filtering Using AS-ST vs LS-ST Thresholds for Known Mixture Samples

2-Person Mixture Samples (80 samples, 4 genotypes, 5 ratios)

Number of Stutter Position Differences

	n-1	n-1/n+1	n-0.5	n+1
Total	13	20	1	0
CT Allele-Specific Correct	9	20	1	
Filtered (stutter)	2	20	1	
Not-Filtered (allele)	7	0	0	
Promega Avg.+3SD Correct	4	0	0	
Filtered (stutter)	4	0	0	
Not-Filtered (allele)	0	0	0	

3-Person Mixture Samples (40 samples, 2 genotypes, 5 ratios)

Number of Stutter Position Differences

	n-1	n-1/n+1	n-0.5	n+1
Total	31	24	1	2
CT Allele-Specific Correct	28	22	1	0
Filtered (stutter)	16	21	1	
Not-Filtered (allele)	12	1	0	
Promega Avg.+3SD Correct	3	2*	0	2
Filtered (stutter)	3	0	0	2
Not-Filtered (allele)	0	2	0	0

AS-ST thresholds produce a more accurate DNA profile especially for stacked stutter positions.
For (80) 2-person mixtures, AS-ST thresholds failed to filter 4 stutter peaks. LS-ST thresholds filtered 7 allele peaks and did not filter 23 stutter peaks.

For (40) 3-person mixtures, AS-ST thresholds failed to filter 5 stutter peaks and filtered 2 allele peaks. LS-ST thresholds filtered 13 allele peaks and did not filter 38 stutter peaks. (* if stacked stutter algorithm used, peaks would be filtered)

3-person Mixture DNA Profile

Contributor Genotypes

	Female	Male-1	Male-2
D3S1358	16	14,16	15,18
D1S1656	12,15	15,16	11,14
D2S441	12,14	11.3,14	13,14
D10S1248	13,15	13,15	13,16
D13S317	10,11	10,12	10,11
Penta E	15,23	13,17	7,12
D16S539	9,12	11,15	12
D18S51	14,20	14,17	14,15
D2S1338	23	17,19	20,24
CSF1PO	11,12	11	10,12
Penta D	9,10	9	10,11
TH01	6,9	9.3	6,9.3
vWA	14,16	17,18	16
D21S11	29,30	29	29,33
D7S820	9.1,11	10,12	9
D5S818	7,13	11,12	11
TPOX	8,11	8	11
D8S1179	12,15	12	12,13
D12S391	19,20	20,22	20
D19S433	13	13,15	13.2,14
D2 33	25.2,28.2	16,32	15,16
D22S1045	11,17	15,17	15,16
D7S391	-	11	10
FGA	17,24	21,22	21,25
D3S576	-	17	17
D5S270	-	18	20

Female:Male-1:Male-2 mixture with 10:4:1 ratio and 500pg of total DNA amplified with PowerPlex® Fusion 6C at 29 cycles and injected on a 3130xl (Applied Biosystems) upgraded for 6-dye chemistries at 3kV, 20seconds.

DNA profile analyzed at 50 RFU and filtered with AS-ST thresholds using GeneMarker®HID

Full detection of Female & Male-1, Allele drop-out for Male-2 at D2S441, D18S51, D2S1338, CSF1PO, and D22S1045

Same DNA profile without stutter filters applied is used to export allele call, height & size for STRmix. Stutter not modeled by STRmix (N-0.5) and non-stutter artifacts (e.g. pull-up) need to be edited out of profile.

STRmix outputs probability-weighted genotype combinations at each locus. Software can compare exemplar DNA profiles and compute likelihood ratios (LR) for each exemplar of the prosecution's hypothesis relative to the defense's hypothesis, typically the probability that the exemplar is a contributor divided by the probability that it is an unknown, unrelated contributor.

LRs for True Contributors	
Female	7.2×10^{35}
Male-1	5.6×10^{22}
Male-2	1.6×10^{14}

LRs for 49 Non-contributors	
43 of 49	0*
Average	5.5×10^{-20}
Max	2.3×10^{-19}
Min	3.1×10^{-29}

* If the genotype of a contributor at any one locus is given a zero probability, then the LR = 0

Similar 3-person mixture of 10:4:1, but Male-1 is degraded with Degradation Index of ~6 (Quantifiler® Trio, Applied Biosystems)

Full detection of Female, Allele drop-out for Male-1 at Penta E, D2S1338, D5S818, and D22S1045 Allele drop-out for Male-2 at D18S51, D2S1338, Penta D, D8SS1179, D19S433, D22S1045, and FGA

LRs for True Contributors	
Female	2.6×10^{42}
Male-1	9.3×10^{16}
Male-2	1.8×10^{8}

LRs for 49 Non-contributors	
Average	2.2×10^{-2}
Max	0.496
Min	2.3×10^{-10}

Questions?
Email: John.Schienman@CT.gov