The Single Most Polymorphic STR Locus:
SE33 Performance in U.S. Populations

http://www.cstl.nist.gov/biotech/strbase
http://www.dna.gov/research/nist
john.butler@nist.gov

Buenos Aires, Argentina September 2009

P-019

John M. Butler[1], Carolyn R. Hill[1], Margaret C. Kline[1], David L. Duewer[1], Cynthia J. Sprecher[2], Robert S. McLaren[2], Dawn R. Rabbach[2], Benjamin E. Krenke[2], and Douglas R. Storts[2]

[1] U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8312 USA
[2] Promega Corporation, Madison, WI 53711 USA

The STR locus SE33 (ACTBP2) located on chromosome 6 (6q14) is arguably the most polymorphic marker examined thus far by the forensic community with a heterozygosity of >0.90 in some populations. Three different primer sets were utilized in this study in order to assess the possibilities of primer binding site mutations. Population variation was measured in 460 U.S. Caucasian, 445 African American, 336 Hispanic, and 202 Asian samples along with mutation rates from almost 400 father-son pairs. In addition, the 10 genomic DNA components in NIST Standard Reference Material SRM 2391b were sequenced and found to exhibit a variety of additional base changes, insertions, and deletions outside of the SE33 repeat region.

DNA Samples

DNA was extracted from anonymous, self-identified samples obtained from two commercial blood banks (Interstate Blood Bank, Memphis, TN and Millennium Biotech, Ft. Lauderdale, FL). A total of 400 father-son sample pairs were provided by DNA Diagnostics (Fairfield, OH) in the form of buccal swabs that were extracted via DNA IQ™ (Promega Corporation, Madison, WI). These samples have been previously typed with autosomal STR [2] and Y-chromosome STR [3] loci.

Self-Identified Ethnicity	#Samples	#Alleles
Caucasian	460	920
African American	445	890
Hispanic	336	672
Asian	202	404
Total	**1443**	2886

Number of Distinguishable Alleles Observed in 1443 Samples

SE33	FGA	D21S11	D12S391	D18S51	D1S1656	D19S433	D2S441
58	29	28	24	23	17	16	15

D2S1338	D10S1248	D22S1045	D3S1358	D8S1179	vWA	D16S539	TH01
13	12	11	11	11	11	9	8

In this study, 58 different SE33 alleles were identified, which is twice the number of the next most variable locus (FGA had 29 alleles). **A total of 343 SE33 genotypes were observed with a heterozygosity of 93.8% across all of the samples examined.**

Mutations Observed in Father-Son Samples

Two SE33 mutations were observed out of 391 father-son sample pairs previously shown to be related [3]. These single-step mutation results – one a gain (20 → 21) and one a loss (15 → 14) – were confirmed with multiple PCR primer sets. The mutation rate of 0.5% (2/391) is similar to that reported with American Association of Blood Banks (AABB) paternity annual report of 0.64% (330/51,940) – see http://www.cstl.nist.gov/biotech/strbase/mutation.htm.

SE33 Allele Frequencies in U.S. Population Groups

The relative percent of each allele in the U.S. Caucasian (Cauc), African American (AfAm), Hispanic (Hisp), and Asian groups are shown below. Observed alleles were sorted by their measured motif (x vs x.2 vs x.3) in order to examine the different distribution patterns.

Allele	Total #	Total %	Populations, % AfAm	Asian	Cauc	Hisp
7	1	0.1	0.1			
8	1	0.1				0.1
11	1	0.1			0.1	
12	11	0.4	0.3		0.5	0.4
13	31	1.1	1.1		1.5	1.0
14	85	2.9	5.1	0.2	2.5	2.4
15	102	3.5	3.9	1.2	3.9	3.9
16	144	5.0	4.8	4.7	4	6.7
17	205	7.1	9.3	4.0	6.2	7.3
18	**268**	9.3	12.1	5.0	7.2	**11.0**
19	249	8.6	**12.2**	6.2	6.6	8.0
20	216	7.5	10.9	9.2	5.4	4.8
21	108	3.7	4.6	6.7	2.4	2.7
22	42	1.5	1.3	1.7	1.5	1.3
23	12	0.4	0.6	1.0	0.2	0.1
24	1	0.1			0.1	
26	1	0.1	0.1			
27	1	0.1				0.1
28	2	0.1	0.1	0.2		
29	1	0.1		0.2		
30	1	0.1				0.1
31	3	0.1	0.1		0.2	
32	1	0.1			0.1	
33	2	0.1			0.1	0.1
34	9	0.3	0.3		0.7	
35	1	0.1	0.1			
36	2	0.1	0.2			

Allele	Total #	Total %	Populations, % AfAm	Asian	Cauc	Hisp
10.2	1	0.1	0.1			
11.2	2	0.1	0.2			
12.2	4	0.1	0.2			0.3
13.2	9	0.3	1.0			
14.2	10	0.3	0.4		0.4	0.3
15.2	8	0.3	0.3			0.7
16.2	5	0.2	0.3		0.1	0.1
17.2	1	0.1	0.1			
20.2	20	0.7	0.3	1.2	1.1	0.3
21.2	48	1.7	1.1	1.7	2.4	1.3
22.2	65	2.3	0.4	3.2	3.8	1.9
23.2	91	3.2	2.2	4.2	4.3	2.1
24.2	74	2.6	1.3	6.2	2.2	2.5
25.2	110	3.8	2.7	6.9	4	3.1
26.2	163	5.6	6.1	5.2	4.3	7.1
27.2	225	7.8	4.3	**10.4**	**9.5**	8.6
28.2	180	6.2	4.4	7.9	7.4	6.1
29.2	147	5.1	2.7	5.7	6.3	6.3
30.2	111	3.8	1.6	3.2	5.8	4.6
31.2	52	1.8	1.5	2.5	2.2	1.3
32.2	25	0.9	0.4	0.7	1.3	0.9
33.2	11	0.4	0.3		0.5	0.4
34.2	1	0.1			0.1	

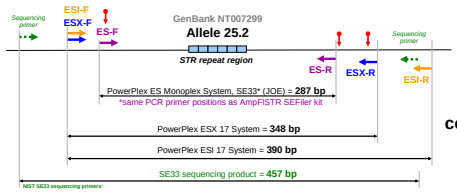
Allele	Total #	Total %	Populations, % AfAm	Asian	Cauc	Hisp
6.3	1	0.1				0.1
16.3	2	0.1				0.3
17.3	5	0.2	0.1		0.2	0.3
18.3	1	0.1			0.1	
27.3	2	0.1				0.3
28.3	2	0.1	0.1		0.1	
29.3	1	0.1		0.2		

Heterozygosity: Total (0.9377), Caucasian (0.9479), Hispanic (0.9318), African American (0.9305), Asian (0.9261)

PCR Primer Sets Compared

SE33 genotyping was performed with three different and non-overlapping primer sets to enable discovery of potential null alleles in this highly polymorphic STR locus. PowerPlex ES Monoplex System, SE33 (JOE) utilizes the original published Polymeropoulos *et al.* (1992) primers [8] that are also contained in Applied Biosystems' SEFiler kit [5]. The new PowerPlex ESX 17 and ESI 17 Systems share the forward primer position but generate PCR products that are 42 bp different in size due to separate reverse primer positions.

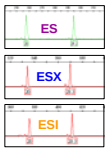
SE33 Relative PCR Primer Positions



99.93% concordance

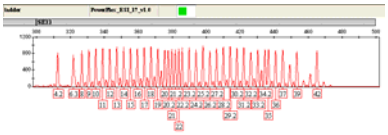
Only Six Discordant Results Were Observed

2886 alleles x 3 primer sets = 8658 comparisons
6/8658 = 0.07% discordance



ES Primers	ESX Primers	ESI Primers	Sequence Reason
26.2, 26.2	26.2, 27.2	26.2, 27.2	C→T 134 bp upstream (impacts ES-F primer)
20,29.2	20,28.3	20,28.3	3 bp deletion (TTG) 41-43 bp downstream (outside ES-R primer)
28.2,28.2	24.2,28.2	24.2,28.2	C→T 134 bp upstream (impacts ES-F primer)
21.2,21.2	21.2,26.2	21.2,26.2	C→T 134 bp upstream (impacts ES-F primer)
24.2,24.2	24.2,25.2	24.2,25.2	C→T 134 bp upstream (impacts ES-F primer)
19,25.2	19,19	19,25.2	C→T 75 bp downstream (impacts ESX-R primer annealing)

PowerPlex ESI 17 Allelic Ladder

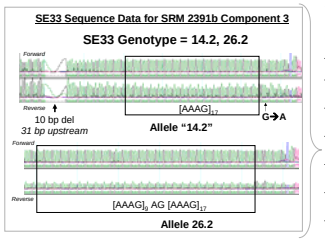


Off-Ladder Alleles

Out of 2286 possible alleles, 27 different off-ladder SE33 alleles were detected a total of 96 times (3.3%).

Allele	Count	Allele	Count	Allele	Count	Allele	Count
7	1	16.2	5	24	1	29.3	1
10.2	1	16.3	2	26	1	30	1
11.2	2	17.2	1	27	1	31	3
12.2	4	17.3	5	27.3	2	32	1
13.2	9	18.3	1	28	2	33	2
14.2	10	19.2	8	28.3	2	34	9
15.2	8	23	12	29	1		

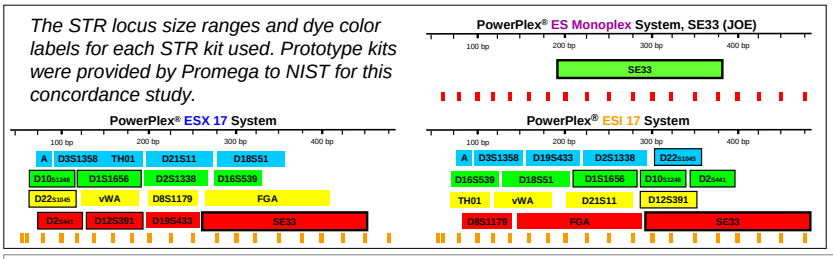
SE33 Genotyping and Sequencing Results for SRM 2391b Components



NIST Standard Reference Material (SRM) 2391b contains certified values for genotyping results and sequencing information on 48 STR markers. SRM 2391b is used by U.S. and international forensic laboratories to help meet Quality Assurance Standards and ISO 17025 traceability requirements. SE33 typing and sequence results are shown below.

Component #	Type	Sequencing Results	Upstream (above repeat)	Downstream (below repeat)
1	20	$[AAAG]_{20}$		
	30.2	$[AAAG]_{13}$ AA AAAG $[AAAG]_{16}$		
2	23.2	$[AAAG]_{7}$ AA $[AAAG]_{16}$		
	28.2	$[AAAG]_{9}$ AA AAAG $[AAAG]_{18}$		
3	"14.2"	$[AAAG]_{17}$	31 bp: 10 bp deletion	4 bp: G->A
	26.2	$[AAAG]_{9}$ AG $[AAAG]_{17}$		
4	"22"	$[AAAG]_{21}$		13bp: AAAG insertion
	28.2	$[AAAG]_{9}$ AA AAAG $[AAAG]_{18}$		
5	14	$[AAAG]_{14}$		
	30.2	$[AAAG]_{13}$ AA AAAG $[AAAG]_{18}$		
6	20	$[AAAG]_{20}$		4 bp: G->A
	21	*unable to get clean sequence*		
7	"13.2"	$[AAAG]_{17}$	11 bp: 14 bp deletion	4 bp: G->A
	20	$[AAAG]_{20}$		4 bp: G->A
8	16	$[AAAG]_{16}$		
	27.2	$[AAAG]_{13}$ AA AAAG $[AAAG]_{14}$		
9	19	$[AAAG]_{19}$		
	29.2	$[AAAG]_{13}$ AA AAAG $[AAAG]_{15}$		
10	23.2	$[AAAG]_{12}$ AA AAAG $[AAAG]_{10}$		
	26.2	$[AAAG]_{11}$ AG $[AAAG]_{15}$		

The STR locus size ranges and dye color labels for each STR kit used. Prototype kits were provided by Promega to NIST for this concordance study.



References

[1] STRBase SE33 Fact Sheet: http://www.cstl.nist.gov/biotech/strbase/str_SE33.htm
[2] Butler, J.M. *et al.* (2003) Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *J. Forensic Sci.* 48(4):908-911.
[3] Decker, A.E. *et al.* (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *Forensic Sci. Int. Genet.* 2(3):e31-e35.
[4] Reid, T.M. *et al.* (2003) Distribution of HUMACTBP2 (SE33) alleles in three North American populations. *J. Forensic Sci.* 48(6): 1422-1423.
[5] Coticone, S.R. *et al.* (2004) Development of the AmpFISTR SEfiler PCR amplification kit: a new multiplex containing the highly discriminating ACTBP2 (SE33) locus. *Int. J. Legal Med.* 118: 224-234.
[6] Hering, S. *et al.* (2002) Sequence variations in the primer binding regions of the highly polymorphic STR system SE33. *Int. J. Legal Med.* 116: 365-367.
[7] Heinrich, M. *et al.* (2004) Allelic drop-out in the STR system ACTBP2 (SE33) as a result of mutations in the primer binding region. *Int. J. Legal Med.* 118: 361-363.
[8] Polymeropoulos, M.H. *et al.* (1992) Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). *Nucleic Acids Res.* 20: 1432.

Funding: Interagency Agreement 2008-DN-R-121 between the National Institute of Justice and NIST Office of Law Enforcement Standards

Poster available for download from STRBase: http://www.cstl.nist.gov/biotech/strbase/pub_pres/Butler-ISFG2009-SE33poster.pdf