The Highly Polymorphic **STR Locus SE33**: History, Characteristics, Concordance, & Population Variation

Carolyn R. (Becky) Hill and John M. Butler
U.S. National Institute of Standards and Technology

February 3, 2011
Promega Webinar

Disclaimer

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

NIST Disclaimer: Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Presentation Outline

- History of SE33 (ACTBP2)
- Locus characteristics
- Allele nomenclature
- Allelic ladders available
- Population variation
- Concordance studies
- Summary (final thoughts)

History of SE33 Use

History of SE33 Use

- 1991, 1992 initial NAR articles (primers defined)
- 1993 FSS examination
- 1993-95 FBI and AFDIL exploration
  - found to be too complex and challenging for the DNA separation systems of the time
- 1993-1997 Brinkmann lab
  - Population studies, nomenclature
- 1994 EDNAP study
- 1998 German DNA database adoption
- 2001-2002 STR kits become available
  - PowerPlex ES (Promega), SEfiler (ABI)
- 2009-2010 Next generation STR kits
  - ESI/ESX 17 (Sept 2009), NGM SElect (early 2011)

First Papers on SE33 (ACTBP2)

6980 Nucleic Acids Research, Vol. 19, No. 24

Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene 2 (ACTBP2) detected using the polymerase chain reaction

D.Warne, C.Watkins, P.Bodfish, K.Nyberg and N.K.Spurr
Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Potters Bar, Herts EN6 3LD, UK

Source/Description: The polymorphic repetitive sequence (AAAG)$_{11}$ AA (AAAG)$_{15}$ is located in the 5′ flanking sequence of human beta-actin pseudogene 2, beginning at base 176 (1). The sequence was identified from a search of the EMBL and GenBank DNA sequence databases (GenBank V00481). The predicted amplified sequence length is 519 bp.

Primer Sequences:
5′GAGAGAGAGAAAGGAAGGAAGG3′
5′ATGGGGTACTTCAGAGTCAGG3′

Frequency:

Allele	Size (bp)	Frequency
A1	519	0.14
A2	513	0.24
A3	501	0.32
A4	485	0.08
A5	467	0.22

Heterozygosity 68% (estimated using 40 unrelated Caucasian individuals)

Mendelian Inheritance: Observed in CEPH families 1029 (Utah pedigree K1345) and 982 (Utah pedigree K1331).

Likely because of poor resolution issues resulting in fewer (unresolved) alleles being observed

First set of primers result in a large sequence length of 519 bp and heterozygosity of 68%

Polymeropoulos et al. (1992) article

1432 Nucleic Acids Research, Vol. 20, No. 6

Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2)

Mihael H.Polymeropoulos, Denise S.Rath, Hong Xiao and Carl R.Merril
National Institute of Mental Health Neuroscience Center, St Elizabeths Hospital, Room 131, 2700 Martin Luther King Avenue, Washington, DC 20032, USA

Chromosomal Localization: We have tentatively assigned the human beta-actin related pseudogene H-beta-Ac-psi-2 to chromosome 6 using rodent/human somatic cell hybrids.

Smaller PCR Product Sizes enabled better resolution of closely spaced alleles

Source/Description: The polymorphic $(AAAG)_n$ repeat begins at base pair 176 of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) on chromosome 6 (1). The polymorphism can be typed using the polymerase chain reaction (PCR) as described previously (2). The predicted length of the amplified sequence was 291 bp.

Primer Sequences:
AATCTGGGCGACAAGAGTGA (AAAG strand)
ACATCTCCCCTACCGCTATA (TTTC strand)

Frequency: Estimated from 78 chromosomes of unrelated individuals. Heterozygosity Index = 93%. PIC = 0.93.

Allele (bp)	Frequency	Allele (bp)	Frequency
A1 318	0.01	A12 270	0.03
A2 314	0.04	A13 266	0.01
A3 310	0.05	A14 262	0.04
A4 306	0.10	A15 258	0.14
A5 302	0.09	A16 254	0.06
A6 298	0.09	A17 250	0.02
A7 294	0.03	A18 246	0.04
A8 290	0.04	A19 242	0.05
A9 282	0.03	A20 238	0.05
A10 278	0.03	A21 234	0.01
A11 274	0.04		

Polymeropoulos primers result in a smaller sequence length of 291 bp and heterozygosity of 93%

First work from Brinkmann (German) Lab

First referred to as SE33 instead of ACTBP2 in this manuscript

Int J Leg Med (1993) 105:315–320

International Journal of
Legal Medicine

Forensic validation of the STR systems SE 33 and TC 11

P. Wiegand[1], B. Budowle[2], S. Rand[1], and B. Brinkmann[1]

[1] Institute of Legal Medicine, Westfälische-Wilhelms-Universität, Von-Esmarch-Strasse 86, W-4400 Münster, Germany
[2] Forensic Science Research and Training Centre, FBI Academy, Quantico, VA 22135, USA

Received September 16, 1992 / Received in revised form November 16, 1992

3 different samples **Ladder**

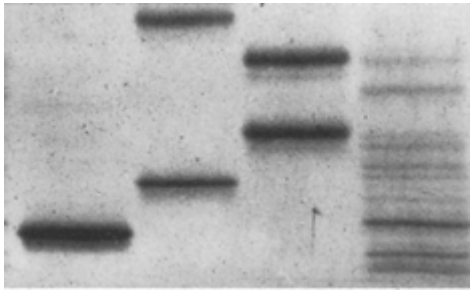


Figure 2b

Poor resolution of alleles is apparent using gel separation systems that were the best available at the time

1994 EDNAP Study

Forensic Science International
65 (1994) 51–59

Report of the European DNA profiling group (EDNAP) — towards standardisation of short tandem repeat (STR) loci

P. Gill*[a], C. Kimpton[a], E. D'Aloja[b], J.F. Andersen[c], W. Bar[d], B. Brinkmann[e], S. Holgersson[f], V. Johnsson[g], A.D. Kloosterman[h], M.V. Lareu[i], L. Nellemann[j], H. Pfitzinger[k], C.P. Phillips[l], H. Schmitter[m], P.M. Schneider[n], M. Stenersen[o]

TH01 was determined to be a suitable candidate for an STR locus, but further work was necessary for SE33 because of difficulties encountered with the reproducibility of migration rates in different electrophoretic systems.

Locus Characteristics

SE33 Locus Characteristics

- **Location:** 6q14 (Chr 6; 89.043 Mb) – beta-actin-related pseudogene
- **Repeat motif**: primarily AAAG (but highly complex patterns)
- **Observed Allele range**: 3 to 49 repeats
- **Heterozygosity**:
  ~ 90-95%
- **Mutation rate**: 0.64%

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 148 (2005) 207–209
Forensic Science International
www.elsevier.com/locate/forsciint

ACTBP2 (alias *ACTBP8*) is localized on chromosome 6 (band 6q14)

S. Wenda[a], E.M. Dauber[a], D.W.M. Schwartz[a], C. Jungbauer[b], V. Weirich[c], R. Wegener[c], W.R. Mayr[a,b,*]

[a]*Division of Blood Group Serology, Medical University Vienna, Waehringer Guertel 18-20, A-1090 Wien, Austria*
[b]*Blood Donation Centre, Austrian Red Cross, Wiedner Hauptstrasse 32, A-1040 Wien, Austria*
[c]*Institute of Legal Medicine, University of Rostock, Sankt-Georg-Strasse 108, D-18055 Rostock, Germany*

Received 5 January 2004; received in revised form 19 May 2004; accepted 21 May 2004
Available online 28 July 2004

Stutter % and Peak Height Ratios will be discussed later

23 STR loci present in STR kits

STR Locus	Alleles Observed	Genotypes Observed	H(obs)	P_I (all samples) n = 1426
SE33	58	341	0.9383	0.0063
Penta E*	20	113	0.8779	0.0175
D2S1338	13	73	0.8752	0.0221
D1S1656	17	99	0.8871	0.0229
D18S51	23	102	0.8696	0.0263
D12S391	24	120	0.8654	0.0279
FGA	29	111	0.8702	0.0299
Penta D*	16	70	0.8733	0.0360
D21S11	32	98	0.8331	0.0399
D19S433	16	83	0.8100	0.0534
D8S1179	11	48	0.7966	0.0553
vWA	11	42	0.8000	0.0624
D16S539	9	30	0.7812	0.0723
D13S317	9	30	0.7749	0.0724
D7S820	12	35	0.7826	0.0745
TH01	9	27	0.7518	0.0752
D2S441	14	46	0.7777	0.0807
D10S1248	12	41	0.7812	0.0828
D3S1358	11	31	0.7489	0.0904
D22S1045	11	45	0.7567	0.0935
D5S818	9	34	0.7225	0.1057
CSF1PO	10	33	0.7567	0.1071
TPOX	10	30	0.6830	0.1351

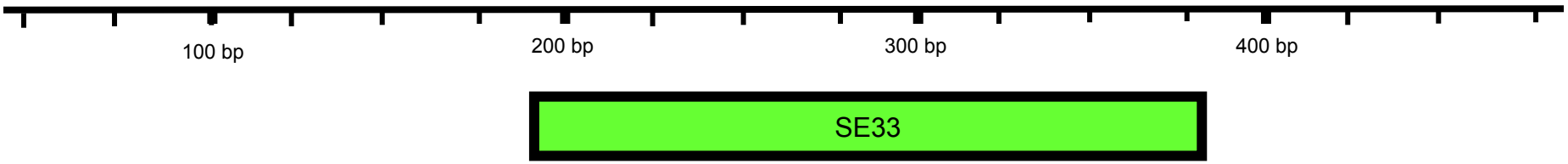
Better for mixtures (more alleles seen)

Better for kinship (low mutation rate)

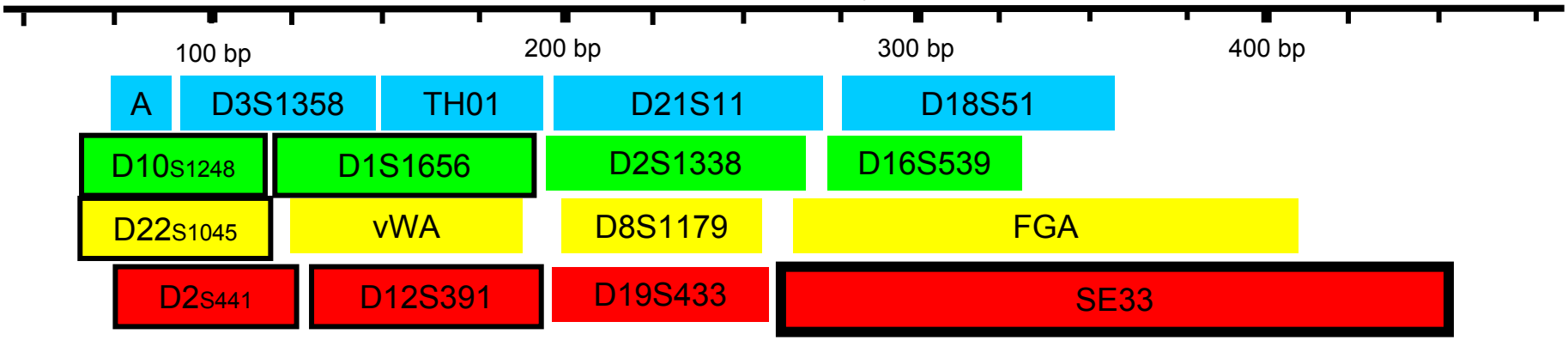
Rank ordered
by their variability

Configurations of STR Kits containing the SE33 Locus

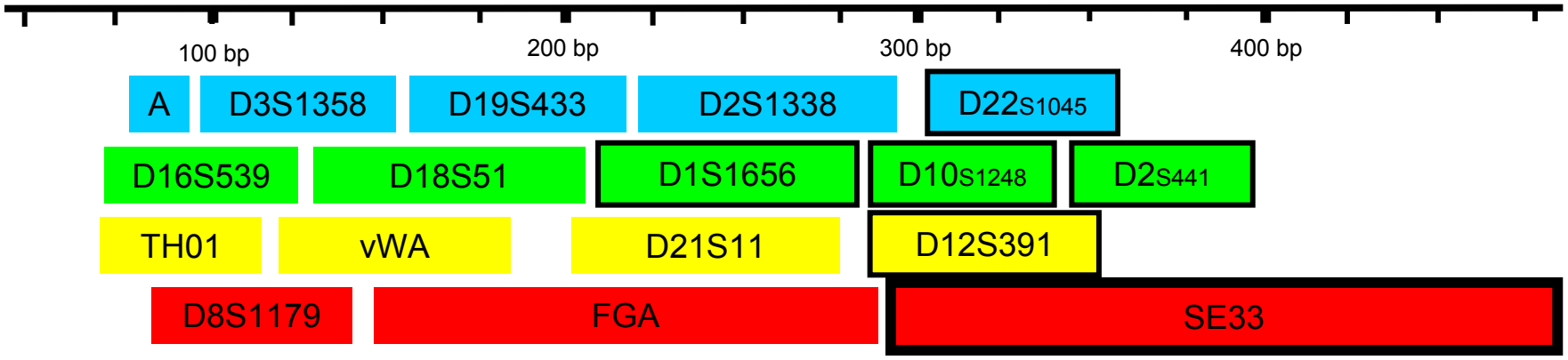
PowerPlex® ES Monoplex System, SE33 (JOE)



PowerPlex® ESX 17 System

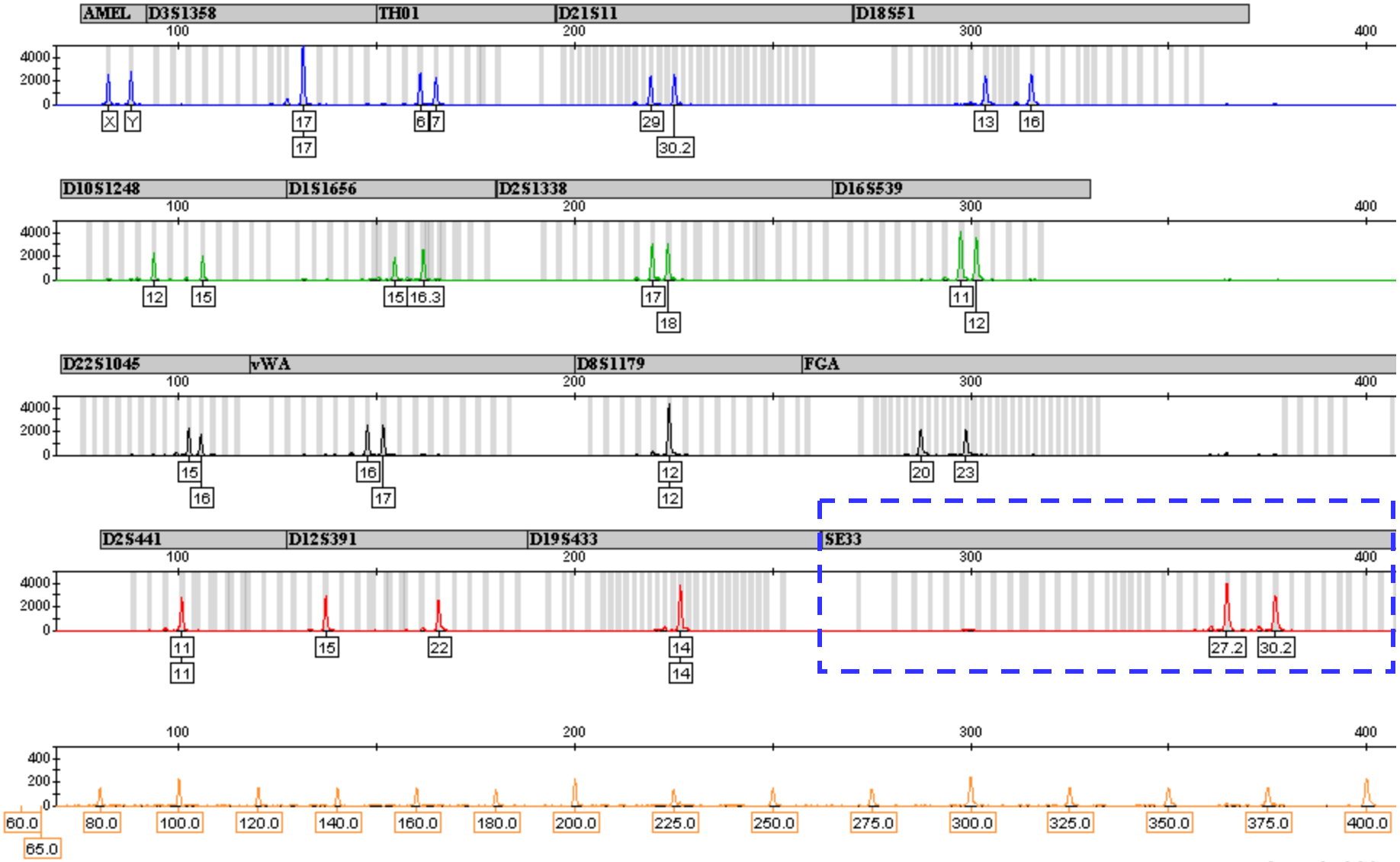


PowerPlex® ESI 17 System



PowerPlex® ESX 17 Example Data

0.5 ng DNA template, 30 cycles



NIST ZT81380

Allele Nomenclature

Allele Nomenclature

Int J Legal Med (1997) 110 : 69–72

ORIGINAL ARTICLE

102 different alleles were observed through sequence analysis

B. Rolf · M. Schürenkamp · A. Junge · B. Brinkmann

Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus

Int J Legal Med (1998) 111 : 97–100

TECHNICAL NOTE

H. R. Schneider · S. Rand · H. Schmitter
G. Weichhold

ACTBP2-nomenclature recommendations of GEDNAP

Important papers that describe SE33 allele nomenclature

171 Published or Known SE33 Alleles

John Butler poster at the 21st International Symposium on Human Identification (San Antonio, TX), October 12-14, 2010, "Characterization of Additional STR Loci for Possible U.S. Core Expansion: D12S391, D1S1656, D2S441, D10S1248, D22S1045, and SE33" [.pdf]

				Repeat Motif Patterns																
Allele (Repeat #)	**ABI SEfiler**	**Promega ESX 17**	**Promega ESI 17**	AAAG	AG	AAAG	AG	**AAAG**	AAAAAG	AG	AGAAAG	AAAG	AAAAAG	AAAG	G	AAGG	AAAG/ANAG	AG	**Reference**
				5' flanking				**central repeat**							3' flanking				
3	197 bp	258 bp	300 bp																STRBase
4.2	203 bp	264 bp	306 bp																PP-ESI ladder
6.3	212 bp	273 bp	315 bp	2	1	3	1	7	0	0	0	0	0	0	0	0	3	1	Rolf *et al.* (1997)
7	213 bp	274 bp	316 bp																Lászik *et al.* (2001)
7.3	216 bp	277 bp	319 bp	2	1	3	1	8	0	0	0	0	0	0	0	0	3	1	Dauber *et al.* (2004)
8	217 bp	278 bp	320 bp																PP-ESI ladder
8.1	218 bp	279 bp	321 bp																Lászik *et al.* (2001)
9 (a)	221 bp	282 bp	324 bp	2	1	3	1	9	0	0	0	0	0	0	1	0	3	1	Dauber *et al.* (2009)
9 (b)	221 bp	282 bp	324 bp	2	1	3	1	9	0	0	0	0	0	0	1	1	2	1	Kline *et al.* (2010)
9.2	223 bp	284 bp	326 bp																Lászik *et al.* (2001)
10	225 bp	286 bp	328 bp																PP-ESI ladder
10.2	227 bp	288 bp	330 bp	2	1	0	0	18	0	0	0	0	0	0	1	0	3	1	Dauber *et al.* (2009)
10.3	228 bp	289 bp	331 bp																Urquhart *et al.* (1993)
11	229 bp	290 bp	332 bp																PP-ESI ladder
11.2	231 bp	292 bp	334 bp	2	1	0	0	15	0	0	0	0	0	0	1	0	3	1	Dauber *et al.* (2004)
12	233 bp	294 bp	336 bp	2	1	3	1	12	0	0	0	0	0	0	1	0	3	1	Rolf *et al.* (1997)
12.2	235 bp	296 bp	338 bp	2	1	3	0	13	0	0	0	0	0	0	1	0	3	1	Rolf *et al.* (1997)
13	237 bp	298 bp	340 bp																PP-ESI ladder
13.2	239 bp	300 bp	342 bp	2	1	3	0	14	0	0	0	0	0	0	1	0	3	1	Rolf *et al.* (1997), Kline *et al.* (2010)
13.3	240 bp	301 bp	343 bp																Poetsch *et al.* (2010)
14 (a)	241 bp	302 bp	344 bp	2	1	3	1	14	0	0	0	0	0	0	1	0	3	1	Rolf *et al.* (1997)
14 (b)	241 bp	302 bp	344 bp	2	1	3	1	14	0	0	0	0	0	0	1	1	2	1	Kline et al. (2010)
14.1	242 bp	303 bp	345 bp																Poetsch *et al.* (2010)
14.2	243 bp	304 bp	346 bp	2	1	3	0	15	0	0	0	0	0	0	1	0	3	1	Kline *et al.* (2010)

SE33 Internal Sequence Variation

Same Length, Different Internal Sequence

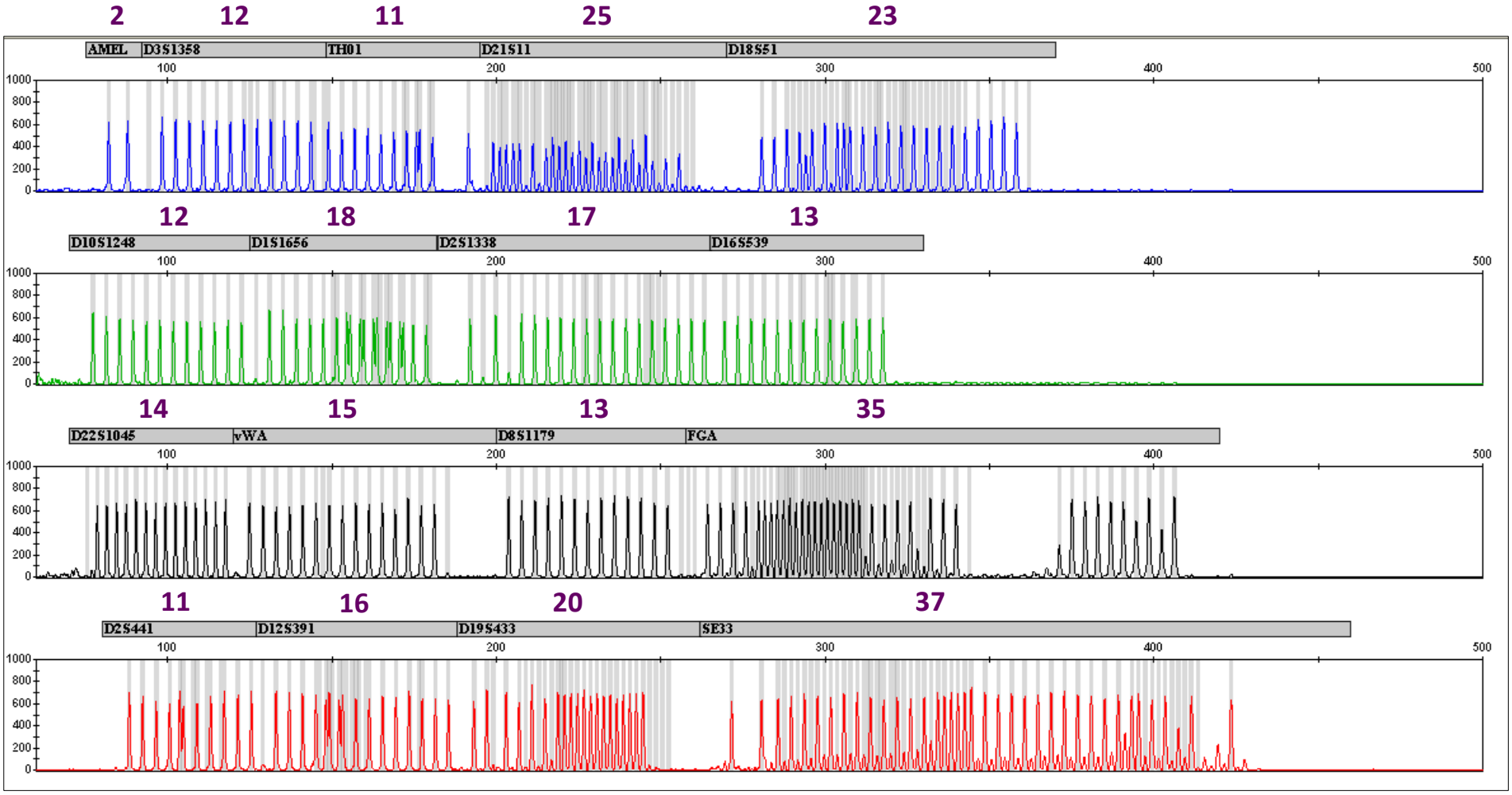
Allele (Repeat #)	ABI SEfiler	Promega ESX 17	Promega ESI 17	Repeat Motif Patterns															Reference
				5' flanking				central repeat							3' flanking				
				AAAG	AG	AAAG	AG	**AAAG**	AAAAAG	AG	AGAAAG	AAAG	AAAAAG	AAAG	G	AAGG	AAAG/ANAG	AG	
28.2 (a)	299 bp	360 bp	402 bp	2	1	3	1	8	1	0	0	19	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (b)	299 bp	360 bp	402 bp	2	1	3	1	9	0	0	0	18	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (c)	299 bp	360 bp	402 bp	2	1	3	1	9	0	0	0	15	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (d)	299 bp	360 bp	402 bp	2	1	3	1	9	1	0	0	18	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (e)				2	1	3	1	10	1	0	0	17	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (f)				2	1	3	1	11	1	0	0	16	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (g)				2	1	3	1	12	1	0	0	15	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (h)	299 bp	360 bp	402 bp	2	1	3	1	13	1	0	0	14	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (i)	299 bp	360 bp	402 bp	2	1	3	1	14	1	0	0	13	0	0	1	1	2	1	Rolf *et al.* (1997)
28.2 (j)	299 bp	360 bp	402 bp	2	1	3	1	14	1	0	0	13	0	0	1	3	0	1	Rolf *et al.* (1997)
28.2 (k)	299 bp	360 bp	402 bp	2	1	3	1	16	1	0	0	11	0	0	1	1	2	1	Rolf *et al.* (1997)
28.3	300 bp	361 bp	403 bp	2	1	3	1	10	1	0	0	12	+A	4	1	1	2	1	Dauber *et al.* (2009)
29	301 bp	362 bp	404 bp	2	1	0	0	15	1	0	0	16	0	0	1	1	2	1	Dauber *et al.* (2009)
29.2 (a)	303 bp	364 bp	406 bp	2	1	3	1	8	1	0	0	20	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (b)	303 bp	364 bp	406 bp	2	1	3	1	9	0	0	1	19	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (c)	303 bp	364 bp	406 bp	2	1	3	1	9	1	0	0	19	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (d)	303 bp	364 bp	406 bp	1	1	3	1	10	1	0	0	19	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (e)	303 bp	364 bp	406 bp	2	1	3	1	11	0	5	0	16	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (f)				1	1	3	1	11	1	0	0	18	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (g)				2	1	3	1	11	1	0	0	17	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (h)				2	1	3	1	12	1	0	0	16	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (i)				2	1	3	1	13	0	0	1	15	0	0	1	3	0	1	Rolf *et al.* (1997)
29.2 (j)	303 bp	364 bp	406 bp	2	1	3	1	13	1	0	0	15	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (k)	303 bp	364 bp	406 bp	2	1	3	1	14	1	0	0	14	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (l)	303 bp	364 bp	406 bp	2	1	3	1	16	1	0	0	12	0	0	1	1	2	1	Rolf *et al.* (1997)
29.2 (m)	303 bp	364 bp	406 bp	2	1	3	1	11	1	0	0	17	0	0	1	1	2	1	D41-TTG-deletion -- Kline *et al.* (2010)

Allele 28.2 (11 sequences)

Allele 29.2 (13 sequences)

Allelic Ladders

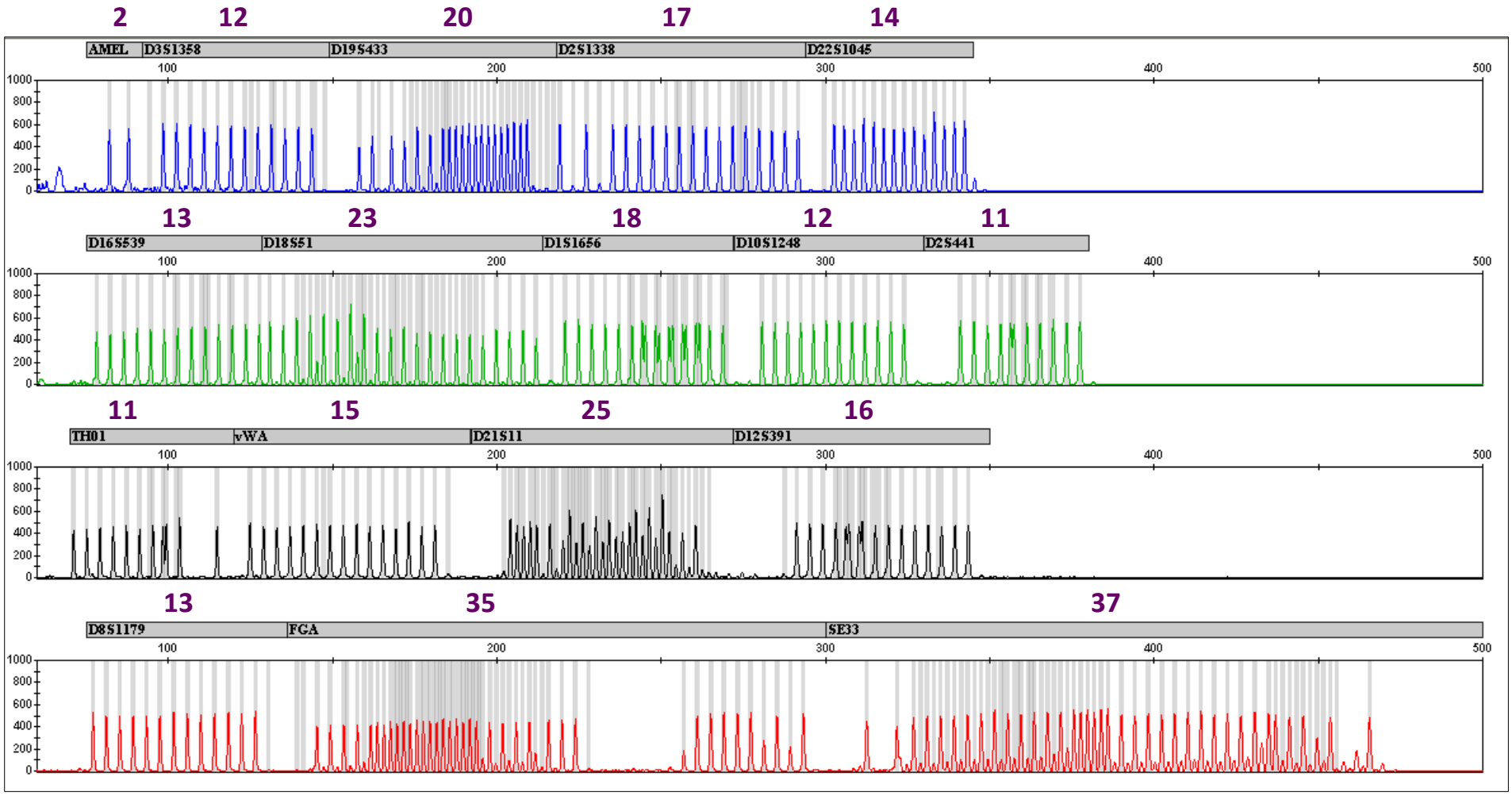
Allelic Ladders for PowerPlex ESX 17



Virtual bins added in for fairly commonly observed micro-variants (i.e. >4 mentions on STRBase)

294 alleles!

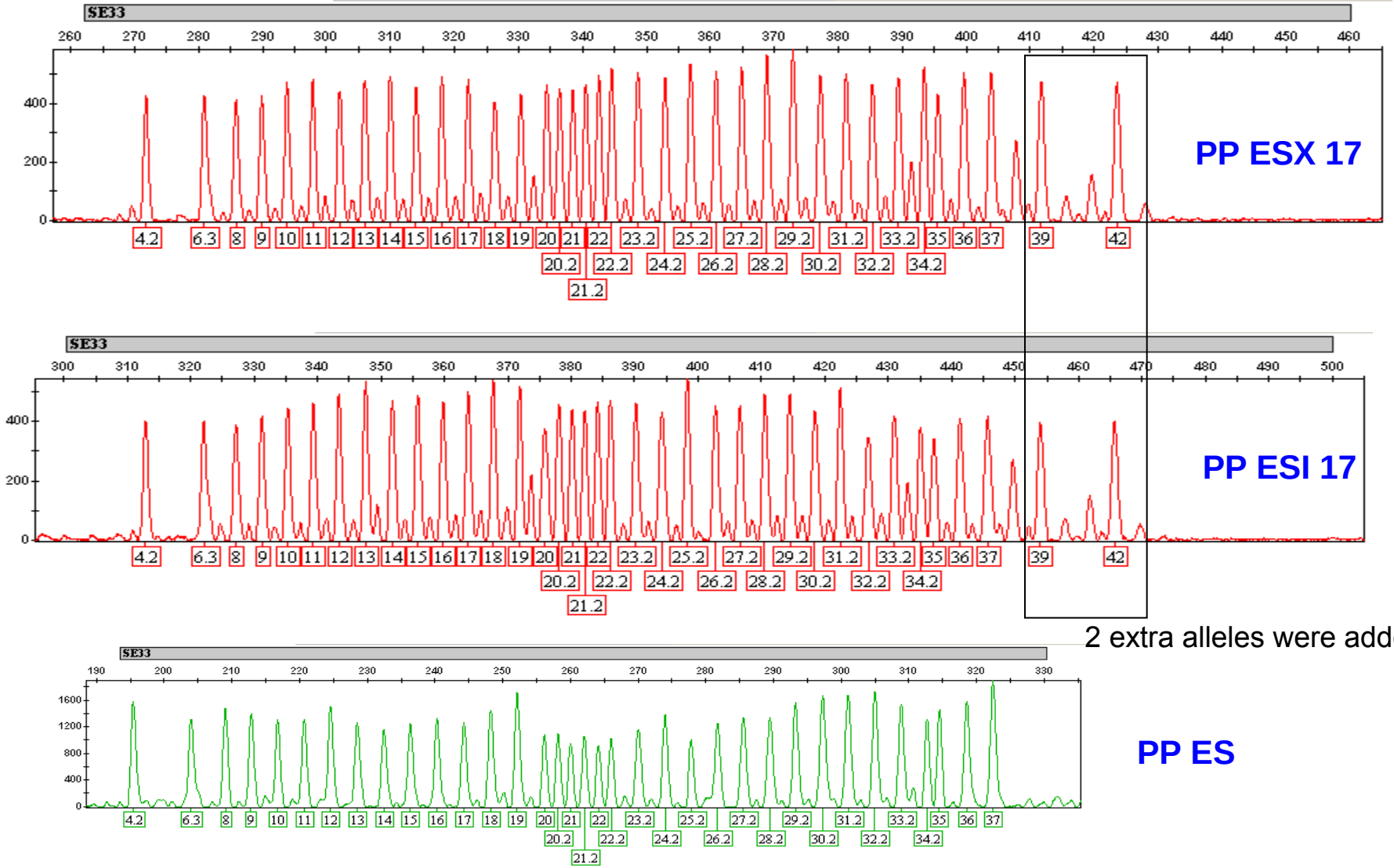
Allelic Ladders for PowerPlex ESI 17



Virtual bins added in for fairly commonly observed micro-variants (i.e. >4 mentions on STRBase)

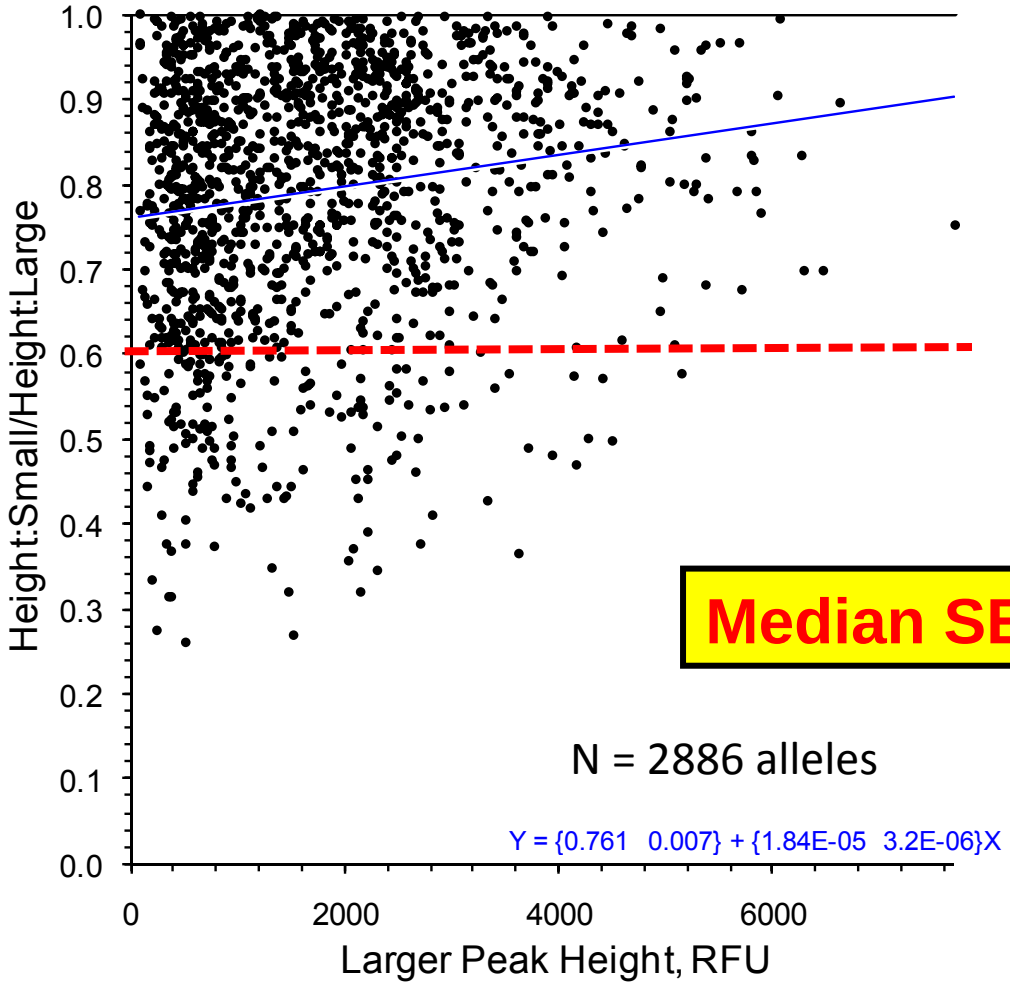
294 alleles!

Comparison of SE33 Ladders



Peak Height Ratios

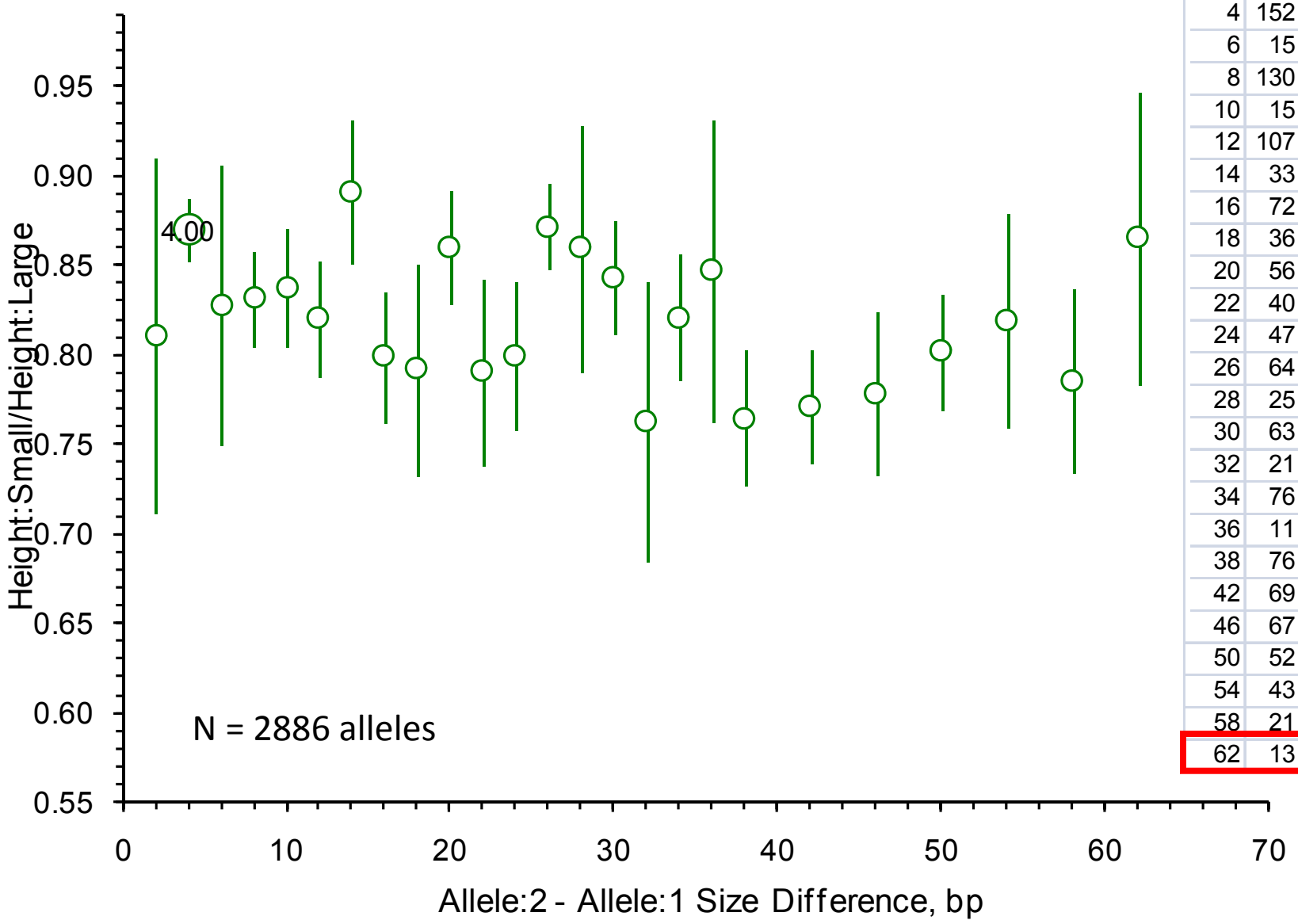
All SE33 PHR Data Points (PP ESI 17)



Median SE33 PHR = 0.896

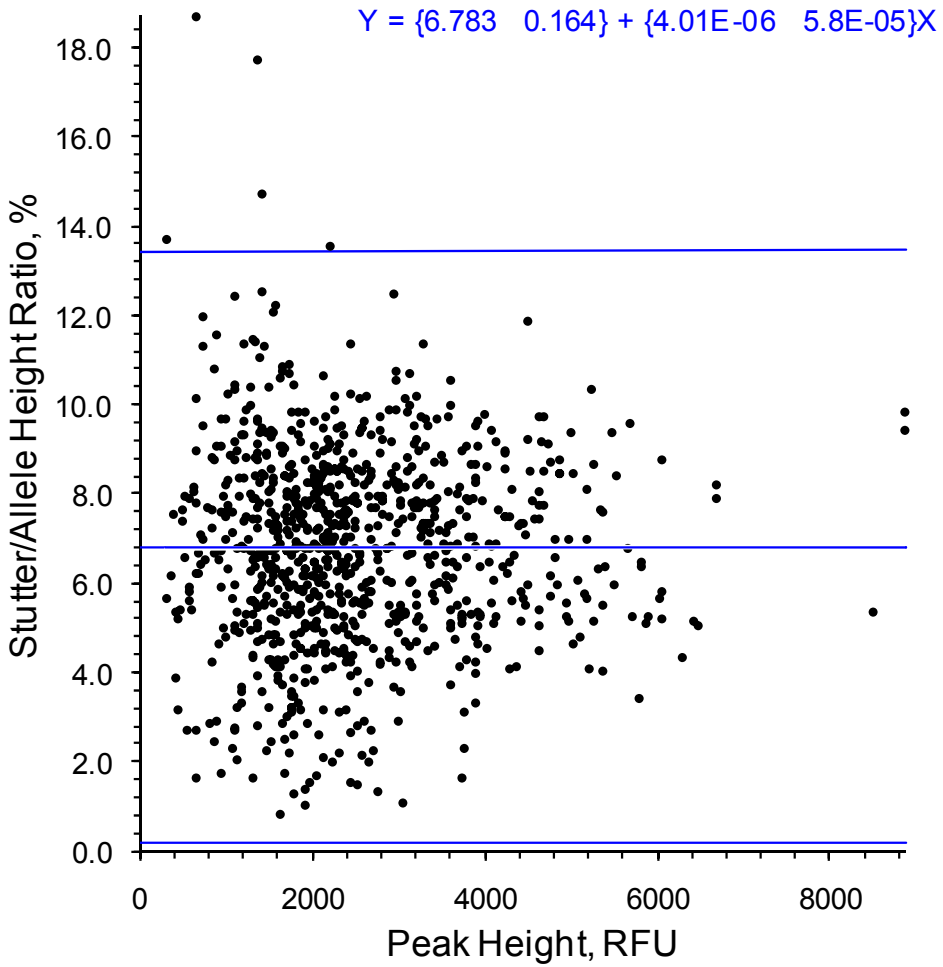
Peak Height Ratios (PP ESI 17)

Between Allele Spreads



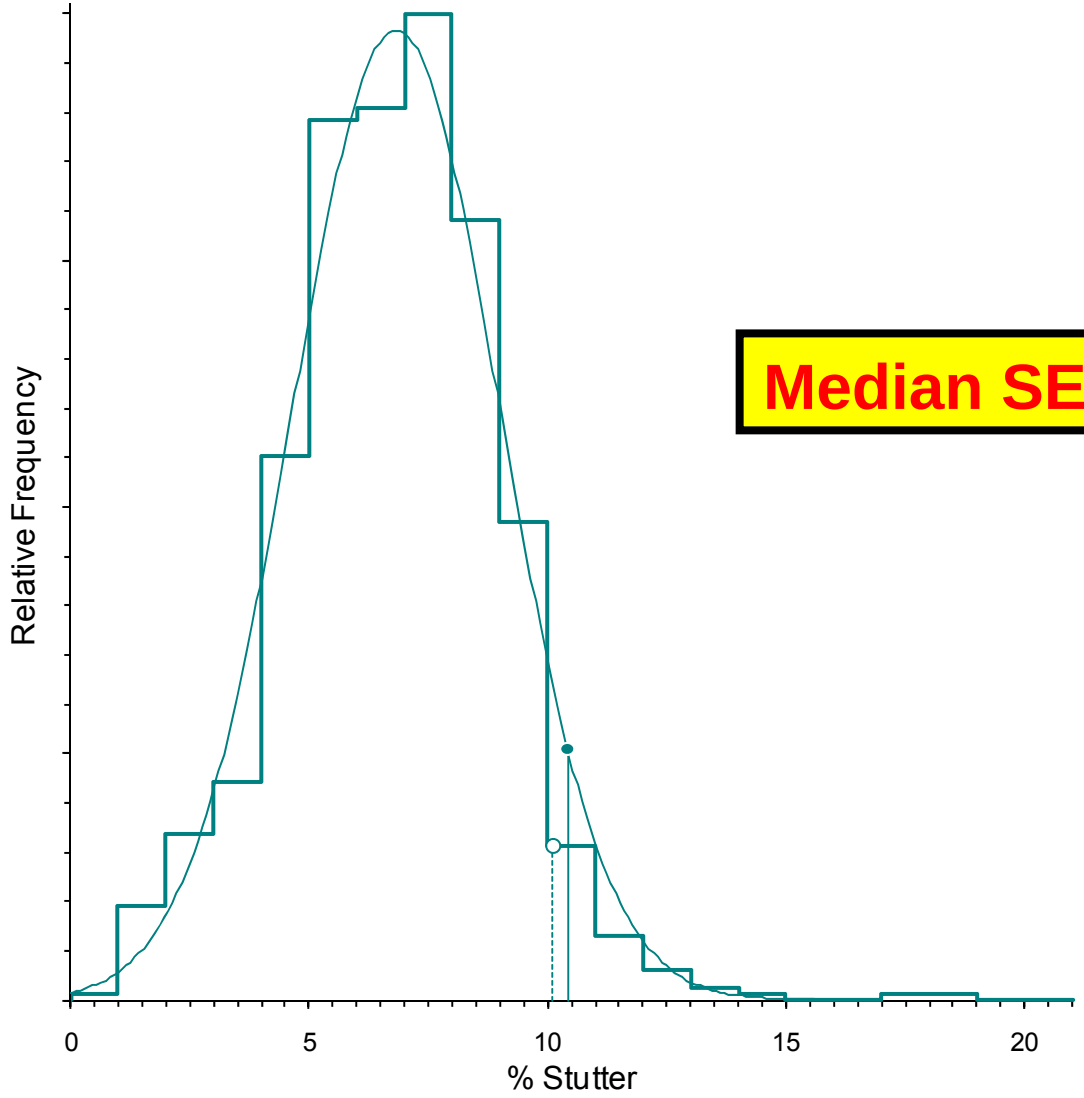
4-bp Motifs			
Δbp	#	Med	U95
2	11	0.81	0.10
4	152	0.87	0.02
6	15	0.83	0.08
8	130	0.83	0.03
10	15	0.84	0.03
12	107	0.82	0.03
14	33	0.89	0.04
16	72	0.80	0.04
18	36	0.79	0.06
20	56	0.86	0.03
22	40	0.79	0.05
24	47	0.80	0.04
26	64	0.87	0.02
28	25	0.86	0.07
30	63	0.84	0.03
32	21	0.76	0.08
34	76	0.82	0.03
36	11	0.85	0.09
38	76	0.77	0.04
42	69	0.77	0.03
46	67	0.78	0.05
50	52	0.80	0.03
54	43	0.82	0.06
58	21	0.79	0.05
62	13	0.87	0.08

SE33 Stutter Percentages



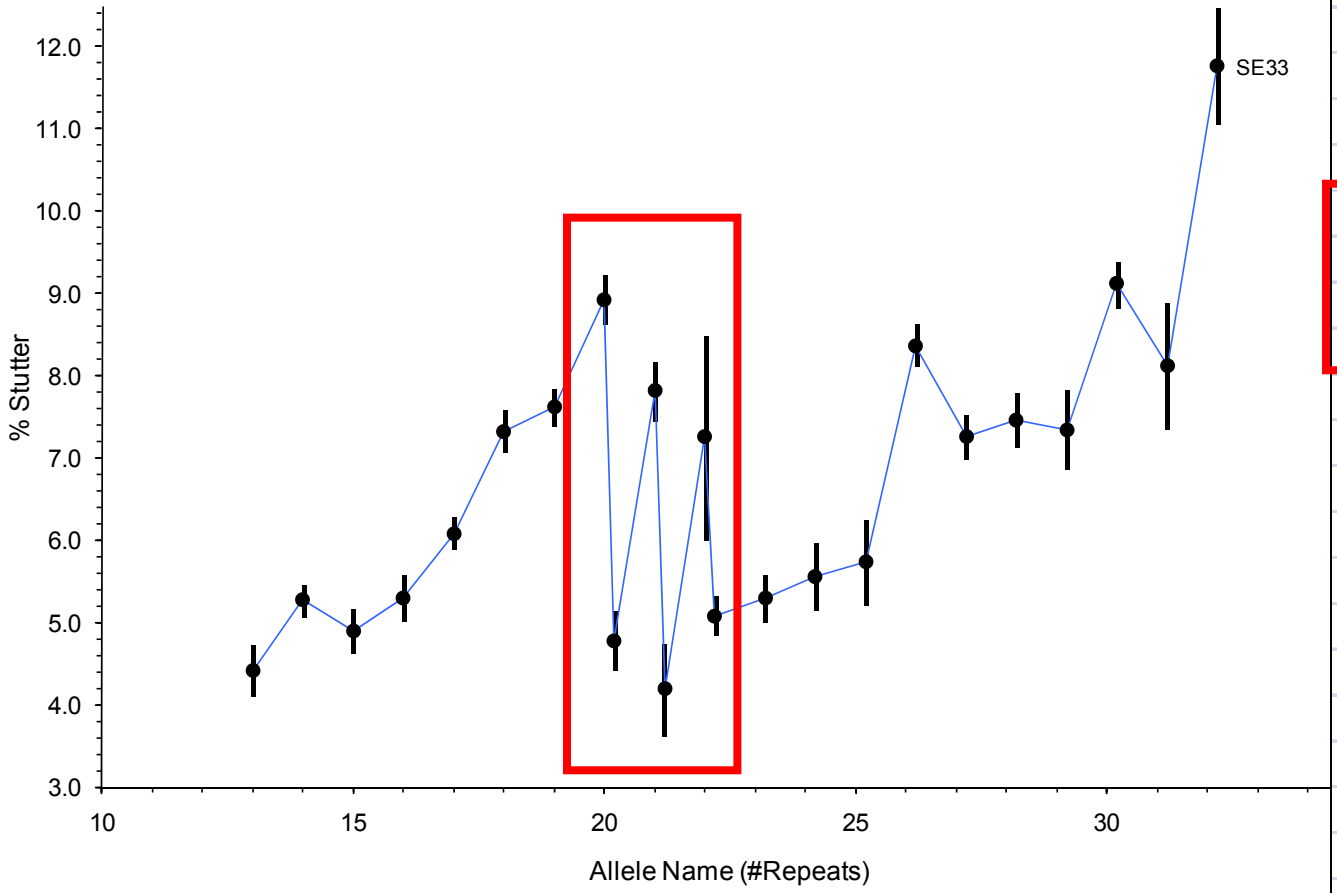
SE33 Stutter Percentages

		Percentage Stutter															
Locus	#Ratio	Mean	SD	Median	MADe	Min	1%	2.5%	5%	10%	25%	75%	90%	95%	97.5%	99%	Max
SE33	901	6.8	2.2	6.8	2.2	0.8	1.6	2.2	3.0	4.1	5.3	8.2	9.4	10.1	10.8	12.1	18.7



Median SE33 Stutter = 6.8%

Stutter Increases with Allele Length



		Stutter		
Allele	Size	#	Median	MADe
13.0	343.3	13	4.4	0.5
14.0	347.5	36	5.3	0.6
15.0	351.5	43	4.9	0.9
16.0	355.6	45	5.3	0.9
17.0	359.6	63	6.1	0.8
18.0	363.6	88	7.3	1.2
19.0	367.6	89	7.6	1.1
20.0	371.6	65	8.9	1.2
20.2	373.6	7	4.8	0.5
21.0	375.5	30	7.8	1.0
21.2	377.8	16	4.2	1.1
22.0	379.8	15	7.2	2.4
22.2	381.8	20	5.1	0.5
23.2	385.9	22	5.3	0.7
24.2	390.0	24	5.6	1.0
25.2	394.1	35	5.7	1.6
26.2	398.1	47	8.4	0.9
27.2	402.3	47	7.3	0.9
28.2	406.2	66	7.5	1.3
29.2	410.3	40	7.3	1.5
30.2	414.2	43	9.1	0.9
31.2	418.3	26	8.1	2.0
32.2	422.3	10	11.8	1.1
	Avg	890	6.7	1.2
	SD		1.8	

Population Variation

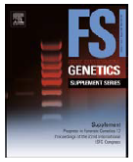
SE33 Variation in U.S. Populations

Forensic Science International: Genetics Supplement Series 2 (2009) 23–24

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS



Research article

The single most polymorphic STR Locus: SE33 performance in U.S. populations

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

[a] National Institute of Standards and Technology, Gaithersburg, MD 20899-8312, USA
[b] Promega Corporation, Madison, WI 53711, USA

G Model
FSIGEN-590; No. of Pages 7

ARTICLE IN PRESS

Forensic Science International: Genetics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems

Carolyn R. Hill[a,*], David L. Duewer[a], Margaret C. Kline[a], Cynthia J. Sprecher[b], Robert S. McLaren[b], Dawn R. Rabbach[b], Benjamin E. Krenke[b], Martin G. Ensenberger[b], Patricia M. Fulmer[b], Douglas R. Storts[b], John M. Butler[a]

[a] National Institute of Standards and Technology, Chemical Science and Technology Laboratory, Gaithersburg, MD 20899-8312, USA
[b] Promega Corporation, Madison, WI 53711-5399, USA

NIST DNA Samples Used in Population Study

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

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 and Y-chromosome STR loci.

SE33 – Most Alleles Observed

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.

PowerPlex ESI 17 Population Data (N=1443)

Marker	Number of Alleles	Theoretical Genotypes	Genotypes Observed	Heterozygosity	PIC
Amelogenin	2	3	3	--	--
TH01	8	36	25	0.7479	0.7572
D3S1358	11	66	31	0.7493	0.7305
D22S1045	11	66	45	0.7548	0.7318
D2S441	15	120	47	0.7729	0.7499
D16S539	9	45	30	0.7791	0.7650
D10S1248	12	78	41	0.7805	0.7460
D8S1179	11	66	48	0.7971	0.7961
vWA	11	66	42	0.7999	0.7866
D19S433	16	136	83	0.8089	0.7984
D21S11	28	406	95	0.8296	0.8293
D12S391	24	300	120	0.8650	0.8651
FGA	29	435	111	0.8691	0.8598
D18S51	23	276	103	0.8698	0.8699
D2S1338	13	91	73	0.8726	0.8821
D1S1656	17	153	101	0.8837	0.8806
SE33	**58**	1711	**343**	**0.9377**	0.9426

SE33 Allele Frequencies **(58 alleles observed)**

343 genotypes observed
Heterozygosity = 0.9377

Allele	Total #	Total %	Populations, % Af Am	Asian	Cauc	Hisp
6.3						
7						
8						
10.2						
11						
11.2	2	0.1	0.2			
12	11	0.4	0.3		0.5	0.4
12.2	4	0.1	0.2			0.3
13	31	1.1	1.1		1.5	1.0
13.2	9	0.3	1.0			
14	85	2.9	5.1	0.2	2.5	2.4
14.2	10	0.3	0.4		0.4	0.3
15	102	3.5	3.9	1.2	3.9	3.9
15.2	8	0.3	0.3			0.7
16	144	5.0	4.8	4.7	4.0	6.7
16.2	5	0.2	0.3		0.1	0.1
16.3	2	0.1				0.3
17	205	7.1	9.3	4.0	6.2	7.3
17.2	1	0.0	0.1			
17.3	5	0.2	0.1		0.2	0.3
18	268	9.3	12.1	5.0	7.2	**11.0**
18.3	1	0.0			0.1	
19	250	8.7	**12.4**	6.2	6.6	8.0
19.2	8	0.3		0.2	0.4	0.4
20	216	7.5	10.9	9.2	5.4	4.8
20.2	20	0.7	0.3	1.2	1.1	0.3
21	108	3.7	4.6	6.7	2.4	2.7
21.2	48	1.7	1.1	1.7	2.4	1.3
22	42	1.5	1.3	1.7	1.5	1.3
22.2	65	2.3	0.4	3.2	3.8	1.9
23	12	0.4	0.6	1.0	0.2	0.1
23.2	91	3.2	2.2	4.2	4.3	2.1
24	1	0.0			0.1	
24.2	74	2.6	1.3	6.2	2.2	2.5
25.2	109	3.8	2.6	6.9	4.0	3.1
26	1	0.0	0.1			
26.2	163	5.6	6.1	5.2	4.3	7.1
27	1	0.0				0.1
27.2	225	7.8	4.3	**10.4**	**9.5**	8.6
27.3	2	0.1				0.3
28	2	0.1	0.1	0.2		
28.2	180	6.2	4.4	7.9	7.4	6.1
28.3	2	0.1	0.1		0.1	
29	1	0.0		0.2		
29.2	147	5.1	2.7	5.7	6.3	6.3
29.3	1	0.0		0.2		
30	1	0.0				0.1
30.2	111	3.8	1.6	3.2	5.8	4.6
31	3	0.1	0.1		0.2	
31.2	52	1.8	1.5	2.5	2.2	1.3
32	1	0.0			0.1	
32.2	25	0.9	0.4	0.7	1.3	0.9
33	2	0.1			0.1	0.1
33.2	11	0.4	0.3		0.5	0.4
34	9	0.3	0.3		0.7	
34.2	1	0.0			0.1	
35	1	0.0	0.1			
36	2	0.1	0.2			

Genotype Frequencies

Most common SE33 genotypes

Genotype	Count
17,18	23
18,19	23
17,19	21
18,27.2	21
17,28.2	20
18,20	20
20,27.2	19
18,26.2	18
20,26.2	18
18,28.2	17
19,19	17
19,27.2	17
26.2,27.2	17
19,20	16
19,26.2	15
20,28.2	15
27.2,28.2	15
27.2,29.2	15

SE33 works well for mixture interpretation

129 occur only once

50 occur twice
39 occur three times
18 occur four times
19 occur five times
16 occur six times
12 occur seven times
12 occur eight times
7 occur nine times
8 occur ten times
6 occur eleven times
3 occur twelve times
3 occur thirteen times
4 occur fourteen times

Locus	#Types	PI
SE33	343	0.0063

Mutations Observed in Father-Son Samples

- 391 father-son samples were examined
- 2 SE33 mutations were observed
  - one gain (20 → 21)
  - one loss (15 → 14)
- Mutation rate = 0.5%
- AABB mutation rate = 0.64%
  - 330/51,940

see http://www.cstl.nist.gov/biotech/strbase/mutation.htm

Off-Ladder Alleles for SE33

NIST Population Study

Out of 2886 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		

Butler, J.M., Hill, C.R., Kline, M.C., Duewer, D.L., Sprecher, C.J., McLaren, R.S., Rabbach, D.R., Krenke, B.E., Storts, D.R. (2009) The single most polymorphic STR locus: SE33 performance in U.S. populations. *Forensic Science International: Genetics Supplement Series (Progress in Forensic Genetics 13*) 2: 23-24.

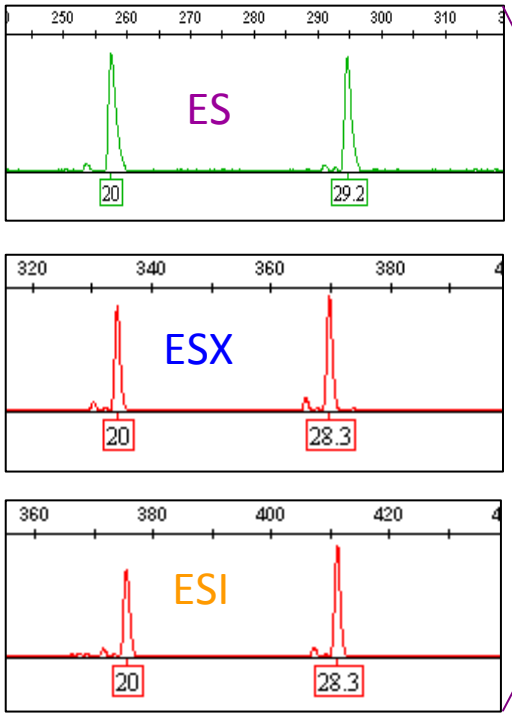
Concordance Studies

Concordance Studies

- Concordance studies are valuable because different primer sets are available
- SE33 primer changes were not an issue really until recently because ABI and Promega used 1992 published primers
- For more information on concordance studies, see Hill *et al* (2010)

Hill, C.R., Kline, M.C., Duewer, D.L., Butler, J.M. (2010) Strategies for concordance testing. *Profiles in DNA (Promega)*, 13(1).

NIST Concordance Results

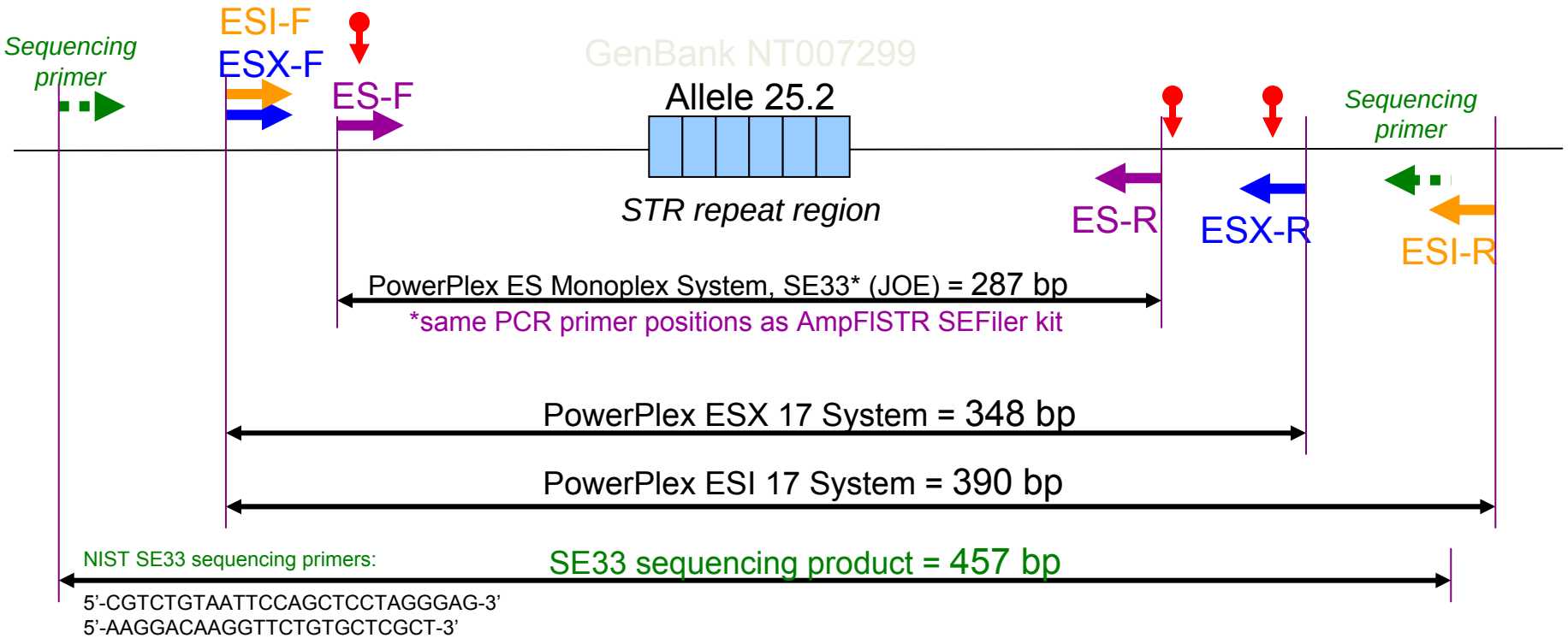


ES Primers	ESX Primers	ESI Primers
26.2, 26.2	26.2, 27.2	26.2, 27.2
20,29.2	20,28.3	20,28.3
28.2,28.2	24.2,28.2	24.2,28.2
21.2,21.2	21.2,26.2	21.2,26.2
24.2,24.2	24.2,25.2	24.2,25.2
19,25.2	19,19	19,25.2

Only Six Discordant Results Were Observed

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

SE33 Relative PCR Primer Positions



99.93%

concordance

Sequence Reasons for Primer Discordance

Sequence Reason
C→T 134 bp upstream (impacts ES-F primer)
3 bp deletion (TTG) 41-43 bp downstream (outside ES-R primer)
C→T 134 bp upstream (impacts ES-F primer)
C→T 134 bp upstream (impacts ES-F primer)
C→T 134 bp upstream (impacts ES-F primer)
C→T 75 bp downstream (impacts ESX-R primer annealing)

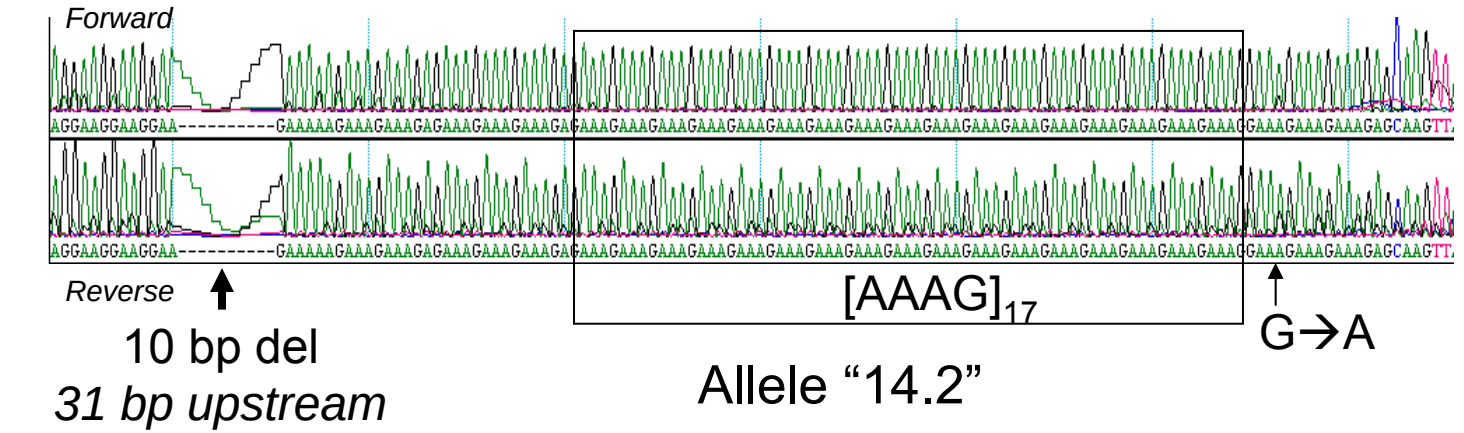
NIST Standard Reference Material (SRM) 2391b contains certified values for genotyping results and sequencing information on 48 STR markers. SRM 2391b is used by U.S. and international forensic laboratories to help meet Quality Assurance Standards and ISO 17025 traceability requirements. SE33 typing and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

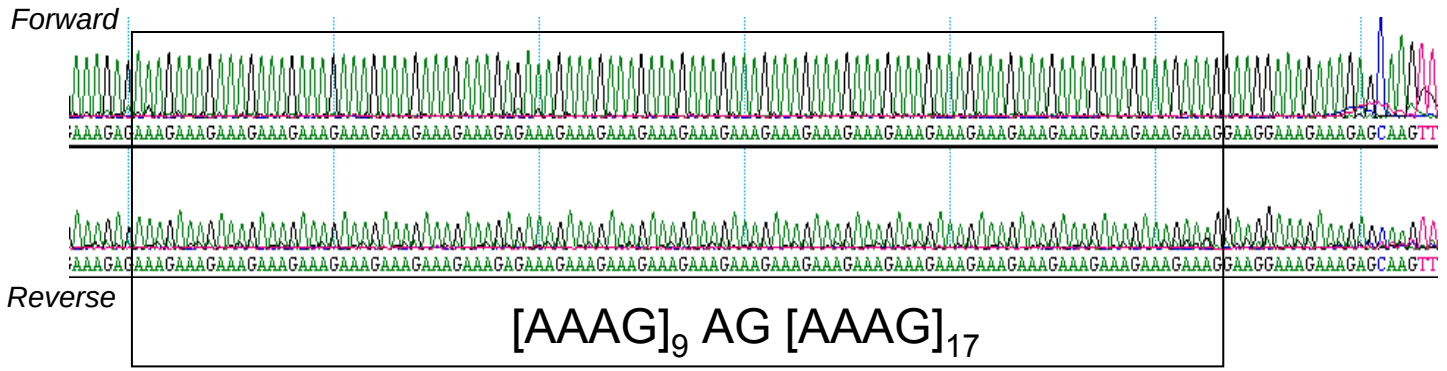
SRM 2391b Certified Values for SE33

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	$[AAAG]_{11}$ 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	$[AAAG]_{12}$ 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 $[AAAG]_{15}$		

SE33 Sequence Data for SRM 2391b Component 3

SE33 Genotype = 14.2, 26.2





ABI NGM SElect Relative Primer Positions

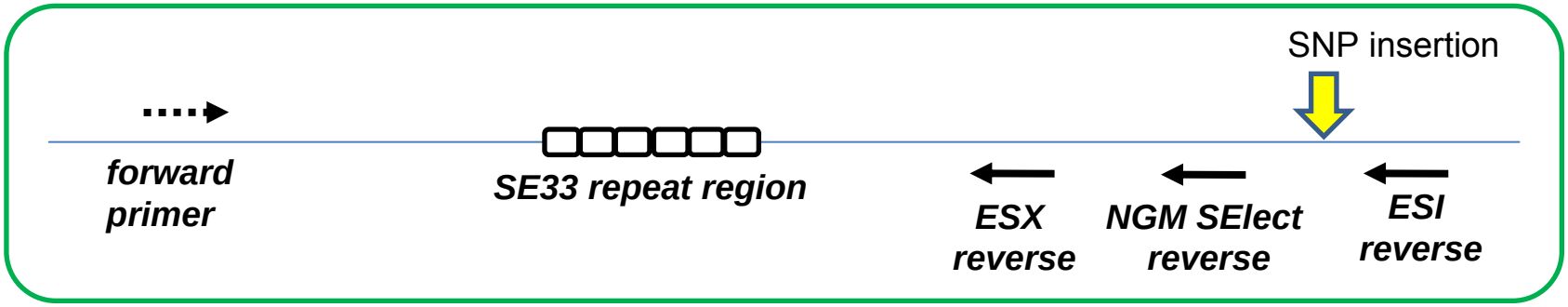
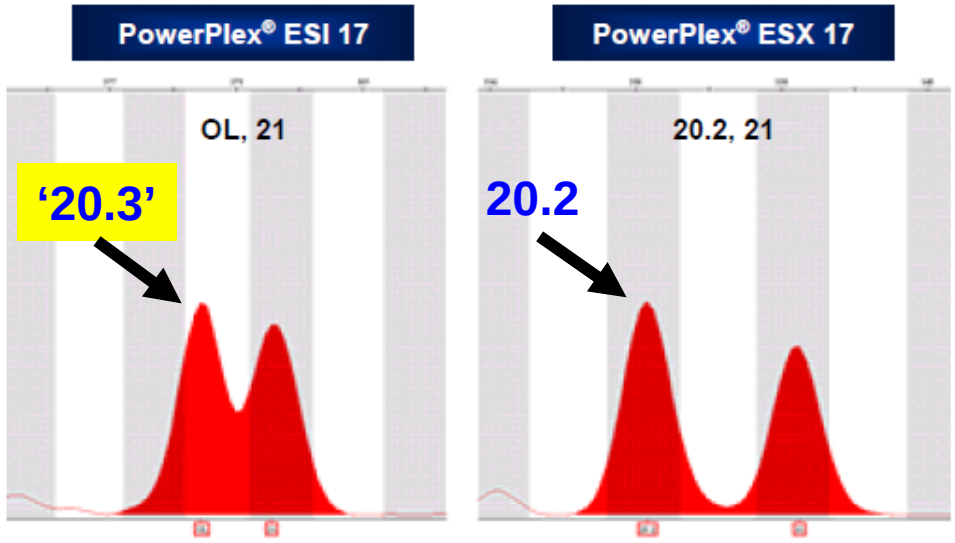


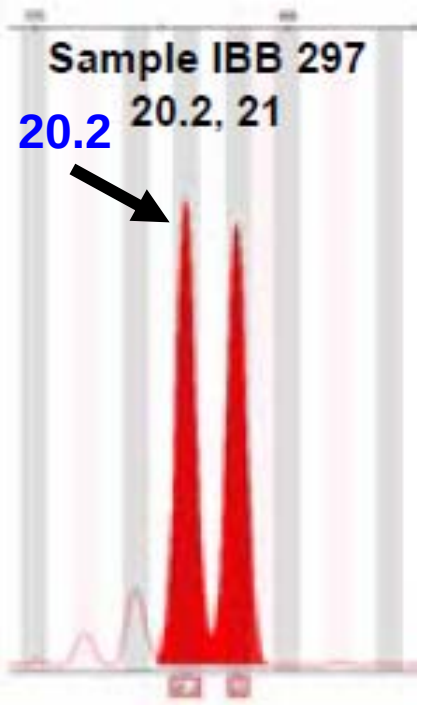
Figure 3. Example of discordance at the SE33 locus for sample IBB 297 between PowerPlex® ESI 17 and ESX 17 results



Each sample which exhibited discordance using the SE33 prototype primers also showed the same discordance when amplified with the ESI kit.

Sequence investigations revealed a SNP-containing region within the prototype SE33 amplicon which, when a SNP occurs, affects the mobility of the amplicon on the capillary electrophoresis platform.

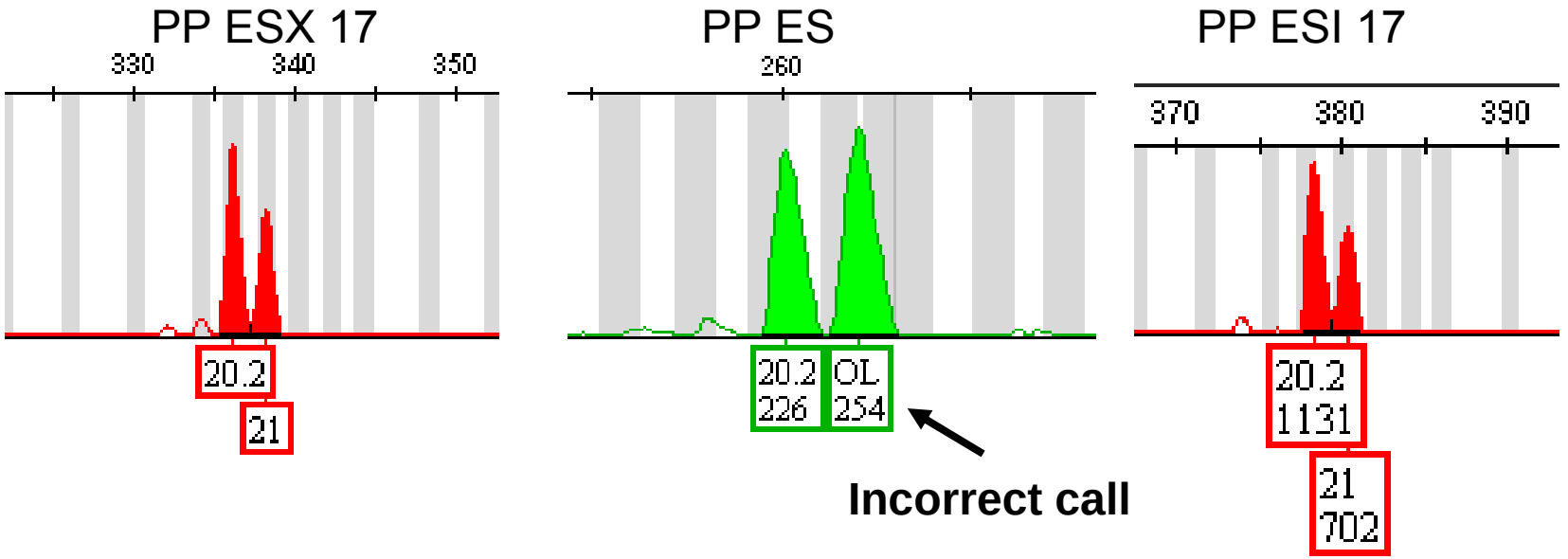
NGM SElect

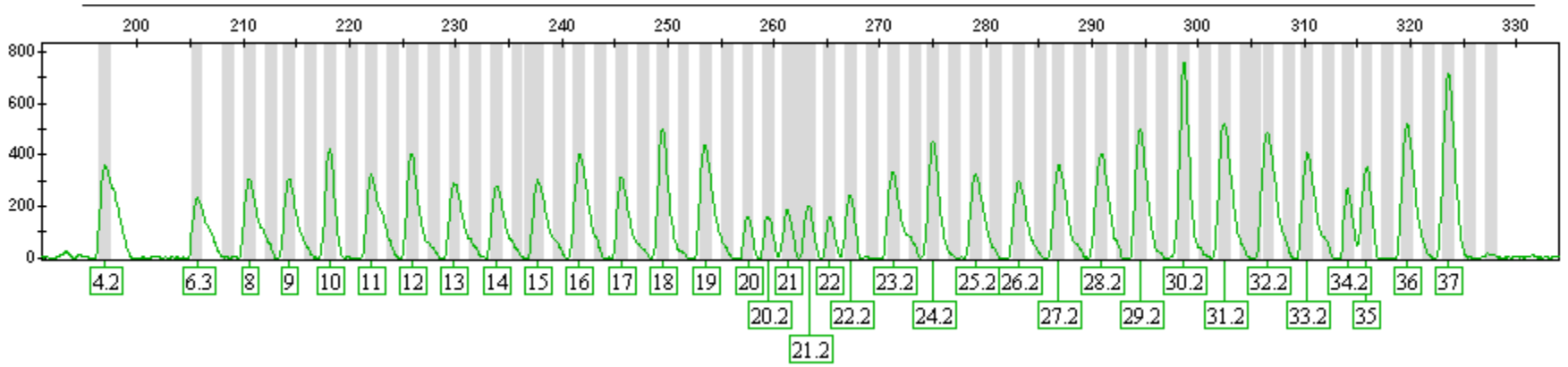


http://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_088773.pdf

Importance of CE Resolution

Differences in CE Resolution Impact Allele Calls





SE33 (PP ES) Ladder with bad resolution

Summary

- SE33 is a powerful marker – especially for mixtures
- Large allele range is problematic for multiplex assays (takes room of ~3 STR loci)
- Between PP ESX 17 and PP ESI 17, we observed one SE33 discordance due to a primer binding site mutation under the ESX reverse primer (No primer sets are completely immune from the possibility of primer binding site mutations)

Acknowledgments

NIST Team for This Work

John Butler
(help w/ slides)

Dave Duewer
(data crunching)

Margaret Kline
(allele sequencing)

A special thanks to Promega for providing the kits used in this study

Promega R&D Scientists

Bob McLaren
Cindy Sprecher
Dawn Rabbach
Ben Krenke
Doug Storts

Becky Hill
301-975-4275
becky.hill@nist.gov