

Email: john.butler@nist.gov
Phone: 301-975-4049

D12S391, D1S1656, D2S441, D10S1248, D22S1045, and SE33

John M. Butler, Carolyn R. Hill, Kristen Lewis O'Connor, David L. Duewer, and Margaret C. Kline

National Institute of Standards and Technology (NIST), 100 Bureau Drive MS 8312, Gaithersburg, MD 20899-8312

In November 2009, the European Union adopted five new autosomal short tandem repeat (STR) loci as part of their expanded European Standard Set (ESS). These new ESS STR loci, which include D12S391 [1], D1S1656 [2], D2S441 [3], D10S1248 [3], and D22S1045 [3], were selected based on discussion over the past few years within the European Network of Forensic Science Institutes (ENFSI) [4,5]. In the past year, Promega Corporation and Applied Biosystems have released new STR kits to enable coverage of these additional loci as well as the highly polymorphic locus SE33 [6,7].

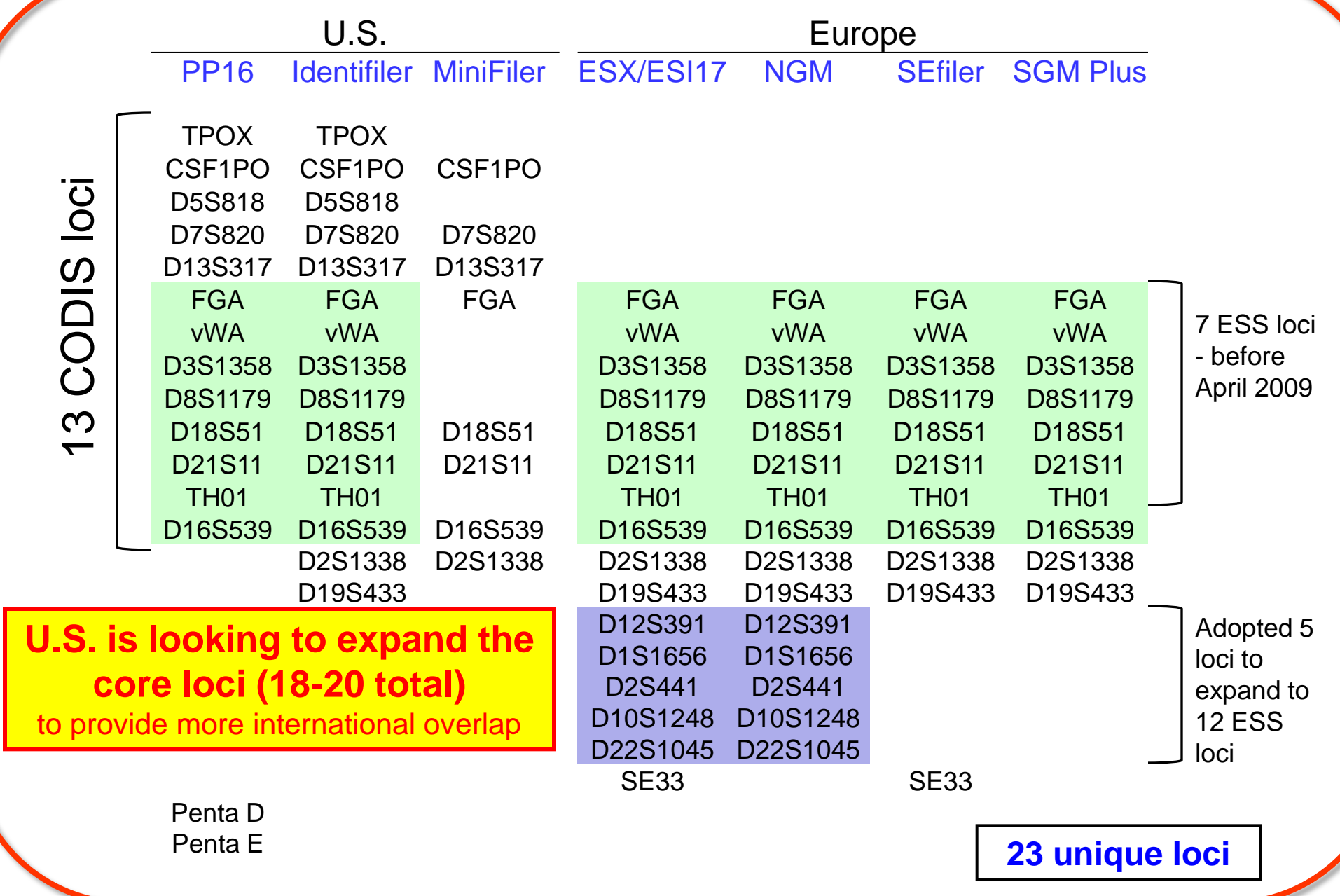
Using three different STR kits (PowerPlex® ESX 17, PowerPlex® ESI 17, and AmpFISTR® NGM), we have studied the allelic variation in over 1440 U.S. population samples [8]. We have also reviewed the literature to find all known variants of these STR loci. Understanding the variation in these additional STR loci across U.S. populations is important because they are being considered as possible candidates for expanding the U.S. core loci in order to enable future international DNA data sharing. Chromosomal location, sequence information, allele frequencies, and power of discrimination are shown for each of these additional autosomal STR loci. In addition, the probability of identity with different sets of loci are illustrated in order to help assess the benefits of adding additional loci to the current 13 CODIS core loci.

Acknowledgments: Funding support from the National Institute of Justice through interagency agreement 2008-DN-R-121 to the NIST Office of Law Enforcement Standards. We thank Promega Corporation for supplying the PowerPlex ES17 and ESX 17 kits and Applied Biosystems for supplying the AmpFISTR NGM kits used for concordance testing purposes.

References

- [1] Lareu, M.V., et al. (1996) A highly variable STR at the D12S391 locus. *Int. J. Legal Med.* 109: 134-138.
- [2] Lareu, M.V., et al. (1998) Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus. *Int. J. Legal Med.* 111: 244-247.
- [3] Coble, M.D. & Butler, J.M. (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. *J. Forensic Sci.* 50: 43-53.
- [4] Gill, P., et al. (2006) The evolution of DNA databases-Recommendations for new European STR loci. *Forensic Sci. Int.* 156: 242-244.
- [5] Gill, P., et al. (2006) New multiplexes for Europe-amendments and clarification of strategic development. *Forensic Sci. Int.* 163: 155-157.
- [6] Moller, A. & Brinkmann, B. (1994) Locus ACTBP2 (SE33): sequencing data reveal considerable polymorphism. *Int. J. Legal Med.* 106: 262-267.
- [7] Butler, J.M., et al. (2009) The single most polymorphic STR locus: SE33 performance in U.S. populations. *Forensic Sci. Int. Genet. Suppl. Ser.* 2: 23-24.
- [8] Hill, C.R., et al. (2010) Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 systems. *Forensic Sci. Int. Genet.* (in press).
- [9] Roll, B., et al. (1997) Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. *Int. J. Legal Med.* 110: 69-72.

Where allele sequence information is available, repeat motif patterns from Roll et al. (1997) have been used as shorthand to indicate the numbers of AAG, AG, or AAAAG repeats that are present



Summary Information

Relative Variability of the 23 STR Loci in Commercial Kits (rank ordered by Probability of Identity (sum of square of observed genotype frequencies))

STR Locus	Alleles Observed	Genotypes Observed	H(obs)	PIC	P _i (all samples) n = 1426	P _i (Cauc) n = 455	P _i (Af Am) n = 439	P _i (Hisp) n = 334	P _i (Asian) n = 198
SE33	58	341	0.9393	0.9424	0.0063	0.0071	0.0104	0.0086	0.0116
Penta E*	20	113	0.8779	0.8992	0.0175	0.0272	0.0200	0.0244	N/A
D2S1338	13	73	0.8752	0.8818	0.0221	0.0280	0.0212	0.0298	0.0334
D1S1656	17	99	0.8871	0.8806	0.0229	0.0206	0.0319	0.0297	0.0444
D18S51	23	102	0.8696	0.8684	0.0263	0.0310	0.0285	0.0304	0.0530
D12S391	24	120	0.8654	0.8646	0.0279	0.0238	0.0366	0.0337	0.0438
FGA	29	111	0.8702	0.8599	0.0299	0.0386	0.0299	0.0271	0.0453
Penta D*	16	70	0.8733	0.8486	0.0360	0.0585	0.0281	0.0529	N/A
D21S11	32	98	0.8331	0.8300	0.0399	0.0489	0.0397	0.0473	0.0558
D19S433	11	83	0.8100	0.7987	0.0534	0.0813	0.0374	0.0619	0.0712
D8S1179	16	48	0.7966	0.7965	0.0553	0.0661	0.0652	0.0634	0.0433
WVA	11	42	0.8000	0.7863	0.0624	0.0696	0.0594	0.0785	0.0811
D16S539	9	30	0.7812	0.7650	0.0723	0.0733	0.0710	0.0771	0.0922
D19S317	9	30	0.7749	0.7637	0.0724	0.0793	0.1317	0.0520	0.0699
D7S820	12	35	0.7826	0.7627	0.0745	0.0626	0.0924	0.0928	0.0922
TH01	9	27	0.7518	0.7578	0.0752	0.0894	0.0999	0.0910	0.1254
D2S441	14	46	0.7777	0.7490	0.0807	0.0867	0.0992	0.1076	0.1020
D10S1248	12	41	0.7812	0.7458	0.0828	0.0975	0.0681	0.1060	0.0873
D3S1358	11	31	0.7489	0.7309	0.0904	0.0726	0.1062	0.0919	0.1355
D22S1045	11	45	0.7567	0.7305	0.0935	0.1253	0.0557	0.1688	0.1070
D5S818	9	34	0.7225	0.7033	0.1057	0.1459	0.0983	0.1317	0.0804
CSF1PO	10	33	0.7567	0.7024	0.1071	0.1327	0.0806	0.1250	0.1120
TPOX	10	30	0.6830	0.6549	0.1351	0.1812	0.0872	0.1525	0.2022

*N=1426 U.S. population samples (Penta D & Penta E from 647 subsets) using PowerPlex 16, ESI 17, and Identifier, data generated at NIST

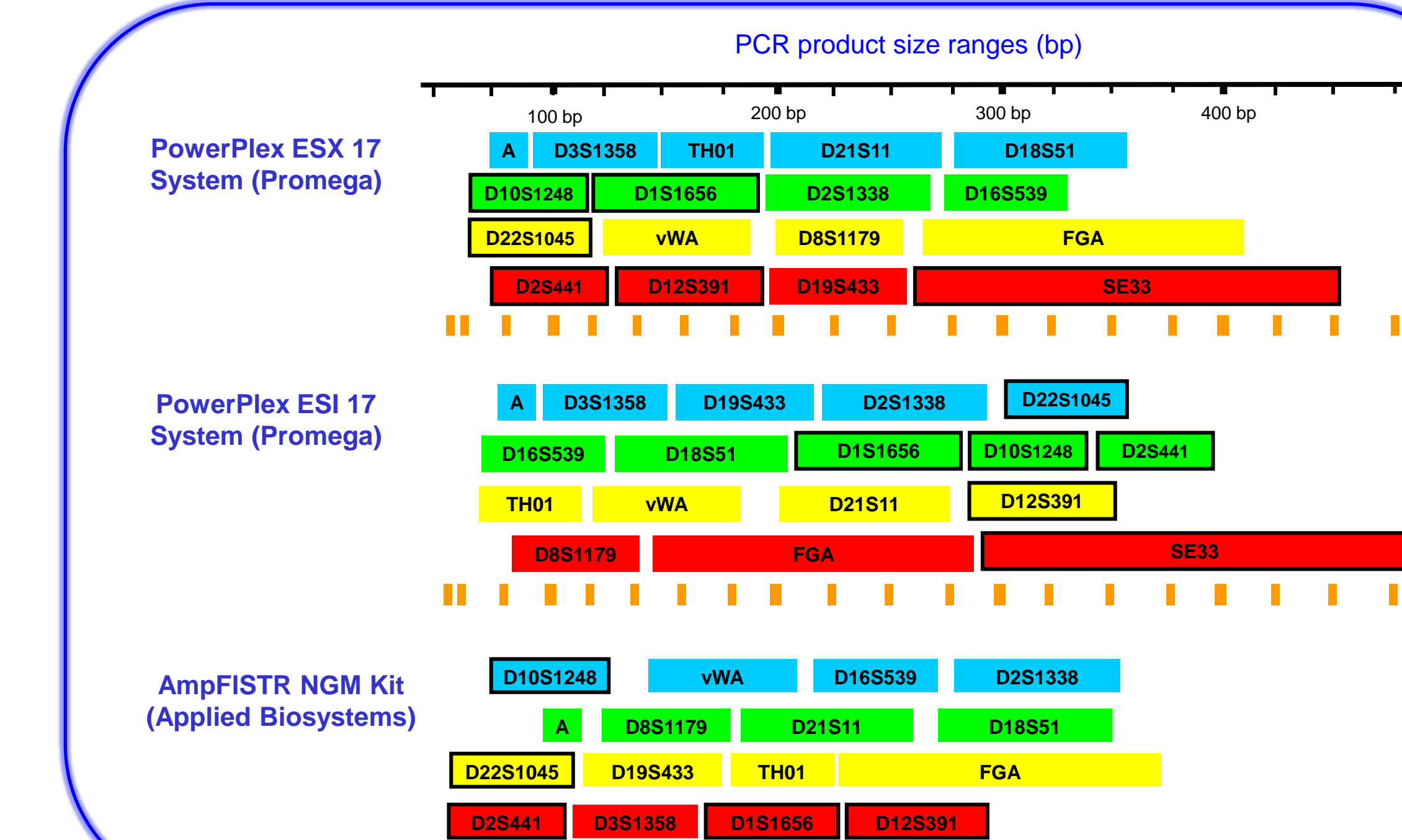
Comparisons of Probability of Identity for Locus Sets

Set of Loci	# STRs	P _i (all samples) n = 1426	P _i (Cauc) n = 455	P _i (Af Am) n = 439	P _i (Hisp) n = 334	P _i (Asian) n = 198
current CODIS	13	4.40E-16	2.95E-15	8.29E-16	1.51E-15	8.65E-15
Identifier	15	5.20E-19	6.72E-18	6.57E-19	2.79E-18	2.05E-17
PowerPlex 16	15	2.77E-19	4.70E-18	4.65E-19	1.95E-18	8.65E-15
NGM	15*	4.04E-19	2.88E-18	5.77E-19	5.64E-18	4.00E-17
ES/ESX 17	16*	2.54E-21	2.05E-20	6.02E-21	4.85E-20	4.66E-19
Extended US1	18	3.25E-22	7.12E-21	2.47E-22	5.36E-21	1.96E-20
Extended US2	19	7.46E-24	1.47E-22	7.87E-24	1.59E-22	8.70E-22
Extended US3	20	4.68E-26	1.05E-24	8.22E-26	1.37E-24	1.01E-23

*VWA removed for statistical calculations due to LD with D12S391

Extended US1 = CODIS 13 + D2S1338, D19S433 + D2S441, D10S1248, D22S1045 + D1S1656
Extended US2 = CODIS 13 + D2S1338, D19S433 + D2S441, D10S1248, D22S1045 + D1S1656
Extended US3 = CODIS 13 + D2S1338, D19S433 + D2S441, D10S1248, D22S1045 + D1S1656 + SE33

Kit Configurations and Concordance Studies



Disconcordance observed during testing 1443 samples

	D12S391	D1S1656	D2S441	D10S1248	D22S1045	SE33
ESX	-	15, 33	9, 1	10/11/12	-	15, 16/17
ESI	-	14, 15, 3	9, 1	10/11/12	-	15, 16/17
NGM	-	14, 15, 3	9, 1	10/11/12	7x	16/17

D2S441 allele 9.1 (occurs primarily in Asian samples) involves the insertion of a G just prior to the TCTA repeat, which disrupts the NGM forward primer annealing. For D22S1045, a G→T mutation 15 bases upstream of the repeat impacts the NGM forward primer (Promega has added a degenerate primer with their ESX kit to overcome this primer binding site mutation).

Disclaimer
Points of view are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. 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.

D12S391

Chr 12: 12.450 Mb

51 observed alleles

Allele (Repeat #)	Promega ESX 17	Promega ESI 17	ABI NGM	Repeat Structure	Reference
14	130 bp	291 bp	250 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
15	134 bp	295 bp	254 bp	(AGAT)AGACAGAT	Lareu et al. (1996)
16	138 bp	299 bp	258 bp	(AGAT)AGACAGAT	Lareu et al. (1996)
17 (a)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Lareu et al. (1996)
17 (b)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (c)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (d)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (e)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (f)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (g)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (h)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (i)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (j)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (k)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (l)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (m)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (n)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (o)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (p)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (q)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (r)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (s)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (t)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (u)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (v)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (w)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (x)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (y)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)
17 (z)	142 bp	303 bp	262 bp	(AGAT)AGACAGAT	Phillips et al. (2006)

D1S1656

Chr 1: 230.905 Mb

25 observed alleles

Allele (Repeat #)	Promega ESX 17	Promega ESI 17	ABI NGM	Repeat Structure	Reference
8	133 bp	225 bp	171 bp	(TAGA)TCT	Phillips et al. (2006)
9	137 bp	229 bp	175 bp	(TAGA)TCT	Phillips et al. (2006)
10 (a)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (b)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (c)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (d)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (e)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (f)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (g)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (h)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (i)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (j)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (k)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (l)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (m)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (n)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (o)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (p)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (q)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (r)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (s)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (t)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (u)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (v)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (w)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (x)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (y)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)
10 (z)	141 bp	233 bp	179 bp	(TAGA)TCT	Phillips et al. (2006)

U.S. Population Data

Hill et al. (2010) FSI Genetics, in press

Allele	#	%	AfAm	Asian	Cauc	Hisp
8	1	0.0	0.2			
9	5	0.2	0.3	1.5	0.2	
10	614	21.3	8.9			