


**American Academy of Forensic Sciences**  
February 24, 2011 – Chicago, IL









## Impact of Additional STR Loci on Random Match Probability Calculations and Kinship Analysis

**John M. Butler,**  
**Carolyn R. Hill, David L. Duewer, Kristen L. O'Connor**

National Institute of Standards and Technology  
Gaithersburg, Maryland





### NIST Human Identity Project Teams within the Applied Genetics Group

<b>Forensic DNA Team</b>				<b>Data Analysis Support</b>	<b>DNA Biometrics Team</b>		
							
John Butler	Mike Coble	Becky Hill	Margaret Kline	Dave Duewer	Pete Vallone	Erica Butts	Kristen Lewis O'Connor

Funding from the National Institute of Justice (NIJ) through NIST Office of Law Enforcement Standards

Funding from the FBI S&T Branch through NIST Information Access Division

### NIST and NIJ Disclaimer

**Funding:** Interagency Agreement between the **National Institute of Justice** and **NIST Office of Law Enforcement Standards**

**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.

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.

### Presentation Outline



- Need for additional loci
  - Growth in U.S. and other national DNA databases
- New STR kits available with additional loci
- Locus characteristics
- Population data and statistical analysis
- Random match probabilities (unrelated)
- Relationship likelihood ratios (related)

AAFS 2009 CRIMINALISTICS Session I:  
*Envisioning the Future*

## Forensic DNA: Perspectives on Progress in a Rapidly Growing Field

John M. Butler, PhD  
National Institute of Standards and Technology

AAFS Meeting (Denver, CO)  
February 19, 2009

Slide from my 2009 AAFS talk...

### Additional STR Loci in the Future?

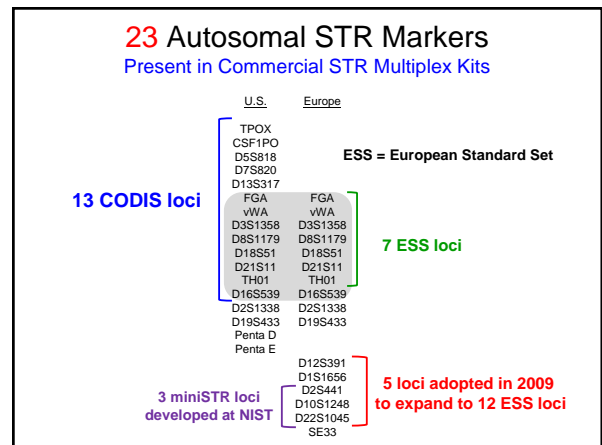
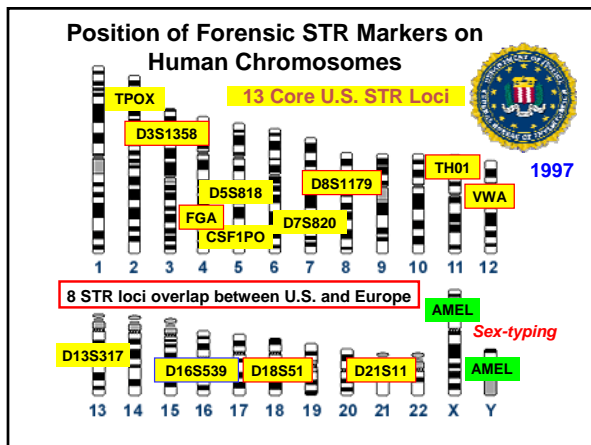
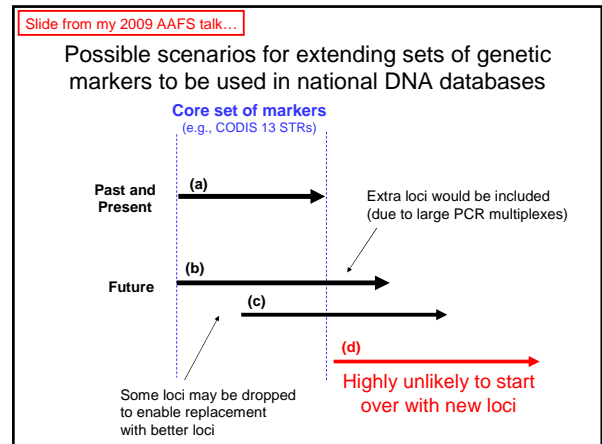
- Will be needed for more complex kinship analyses and extended applications
  - Example: Y-STRs needed for familial searching
- Immigration testing needs more than 13 STRs
- Larger DNA databases will require more loci

### Growth in Numbers of DNA Profiles Present in Various NDIS Indices (cumulative totals by year)

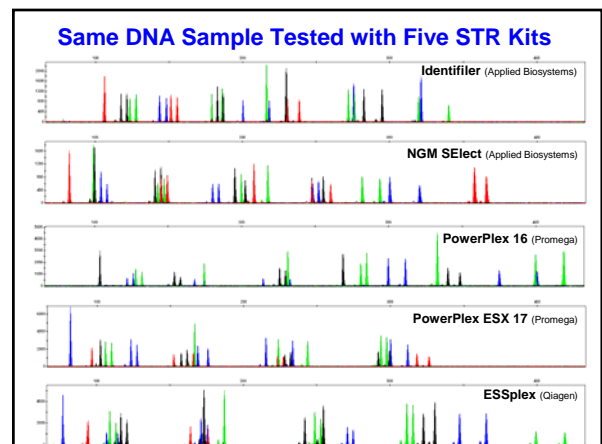
Year ending Dec 31	Forensic	Convicted Offender	Arrestee	Total Offender*
2000	21,625	441,181	--	441,181
2001	27,897	750,929	--	750,929
2002	46,177	1,247,163	--	1,247,163
2003				
2004				
2005				
2006				
2007				
2008	248,943	6,398,874	140,719	6,539,919
2009	298,369	7,389,917	351,926	7,743,329
2010	<b>351,951</b>	<b>8,559,841</b>	<b>668,849</b>	<b>9,233,554</b>

Source: FBI Laboratory's CODIS Unit

In the last two years (2009, 2010):  
**103,008 forensic samples added**  
**2,693,635 offender samples added**



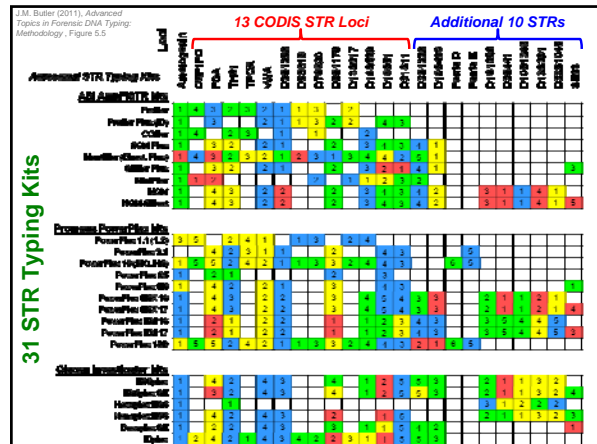
- ### Commercially Available STR Kits
- |  |  |   |
|--|--|---|
| <p><b>Applied Biosystems (17)</b></p> <ul style="list-style-type: none"> <li>-AmpFISTR Blue (1996)</li> <li>-AmpFISTR Green+ (1997)</li> <li>Profiler (1997)</li> <li>Profiler Plus (1997)</li> <li>COfiler (1998)</li> <li>SGM Plus (1999)</li> <li>Identifiler (2001)</li> <li>Profiler Plus ID (2001)</li> <li>-SEfiler (2002)</li> <li>Yfiler (2004)</li> <li>MiniFiler (2007)</li> <li>SEfiler Plus (2007)</li> <li>Sinofiler (2008) - China only</li> <li>Identifiler Direct (2009)</li> <li>NGM (2009)</li> <li>Identifiler Plus (2010)</li> <li>NGM SElect (2010)</li> </ul> | <p><b>Promega Corporation (13)</b></p> <ul style="list-style-type: none"> <li>PowerPlex 1.1 (1997)</li> <li>PowerPlex 1.2 (1998)</li> <li>PowerPlex 2.1 (1999)</li> <li><b>PowerPlex 16</b> (2000)</li> <li>PowerPlex ES (2002)</li> <li>PowerPlex Y (2003)</li> <li>PowerPlex S5 (2007)</li> <li><b>PowerPlex 16 HS</b> (2009)</li> <li>PowerPlex ESX 16 (2009)</li> <li>PowerPlex ESX 17 (2009)</li> <li>PowerPlex ESI 16 (2009)</li> <li>PowerPlex ESI 17 (2009)</li> <li>PowerPlex 18D (2011)</li> </ul> | <p><b>Qiagen (10) kits in 2010</b><br/> <i>Primarily selling kits in Europe<br/>                 Due to patent restrictions cannot sell in U.S.</i></p> <p><b>Investigator kits</b></p> <ul style="list-style-type: none"> <li>ESSplex</li> <li>ESSplex SE</li> <li>Decaplex SE</li> <li>IDplex</li> <li>Nonaplex ESS</li> <li>Hexaplex ESS</li> <li>HDplex</li> <li>Triplex AFS QS</li> <li>Triplex DSF</li> <li>Argus X-12</li> </ul> |
|--|--|---|
- ~1/3 of all STR kits were released in the last two years



### STR Kit Concordance Testing

- Many of these STR kits have different primer sequences for amplifying the same STR locus
- Need to analyze the same DNA samples with different STR typing kits looking for differences
- In some rare cases, allele dropout may occur due to mutations in primer binding regions

See Becky Hill's talk on Friday morning at 10 a.m. (Criminalistics, A110)  
"Concordance Testing Comparing STR Multiplex Kits with a Standard Data Set"



### The 10 STR Loci Beyond the CODIS 13

STR Locus	Location	Repeat Motif	Allele Range*	# Alleles*
D2S1338	2q35	TGCC/TTCC	10 to 31	40
D19S433	19q12	AAGG/TAGG	5.2 to 20	36
Penta D	21q22.3	AAAGA	1.1 to 19	50
Penta E	15q26.2	AAAGA	5 to 32	53
D1S1656	1q42	TAGA	8 to 20.3	25
D12S391	12p13.2	AGAT/AGAC	13 to 27.2	52
D2S441	2p14	TCTA/TCAA	8 to 17	22
D10S1248	10q26.3	GGAA	7 to 19	13
D22S1045	22q12.3	ATT	7 to 20	14
SE33	6q14	AAAG <sup>‡</sup>	3 to 49	178

\*Allele range and number of observed alleles from Appendix 1, J.M. Butler (2011) Advanced Topics in Forensic DNA Typing: Methodology; <sup>‡</sup>SE33 alleles have complex repeat structure

### 25 Alleles Reported in the Literature for D1S1656

Allele (Repeat #)	Promega ESX 17	Promega ESI 17	ABI NGM	Repeat Structure	Reference
8	133 bp	222 bp	171 bp	[TAGA] <sub>8</sub> [TG] <sub>1</sub>	Phillips et al. (2010)
9	137 bp	226 bp	175 bp	[TAGA] <sub>9</sub> [TG] <sub>1</sub>	Phillips et al. (2010)
10 (a)	141 bp	230 bp	179 bp	[TAGA] <sub>10</sub> [TG] <sub>1</sub>	Lareu et al. (1998)
10 (b)	141 bp	230 bp	179 bp	[TAGA] <sub>10</sub> [TAGG][TG] <sub>1</sub>	Phillips et al. (2010)
11	145 bp	234 bp	183 bp	[TAGA] <sub>11</sub> [TG] <sub>1</sub>	Lareu et al. (1998)
12 (a)	149 bp	238 bp	187 bp	[TAGA] <sub>12</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
12 (b)	149 bp	238 bp	187 bp	[TAGA] <sub>12</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
13 (a)	153 bp	242 bp	191 bp	[TAGA] <sub>13</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
13 (b)	153 bp	242 bp	191 bp	[TAGA] <sub>13</sub> [TAGG][TG] <sub>1</sub>	Phillips et al. (2010)
13.3	156 bp	245 bp	194 bp	[TAGA] <sub>13</sub> [TAGA] <sub>1</sub> [TAGG][TG] <sub>1</sub>	Phillips et al. (2010)
14 (a)	157 bp	246 bp	195 bp	[TAGA] <sub>14</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
14 (b)	157 bp	246 bp	195 bp	[TAGA] <sub>14</sub> [TG] <sub>1</sub>	Phillips et al. (2010)
14.3	160 bp	249 bp	198 bp	[TAGA] <sub>14</sub> [TGA] <sub>1</sub> [TAGA] <sub>1</sub> [TAGG][TG] <sub>1</sub>	Phillips et al. (2010)
15	161 bp	250 bp	199 bp	[TAGA] <sub>15</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
15.3	164 bp	253 bp	202 bp	[TAGA] <sub>15</sub> [TGA] <sub>1</sub> [TAGA] <sub>1</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
16	165 bp	254 bp	203 bp	[TAGA] <sub>16</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
16.3	168 bp	257 bp	206 bp	[TAGA] <sub>16</sub> [TGA] <sub>1</sub> [TAGA] <sub>1</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
17	169 bp	258 bp	207 bp	[TAGA] <sub>17</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
17.1	170 bp	259 bp	208 bp	Not published	Schröter et al. (2000)
17.3	172 bp	261 bp	210 bp	[TAGA] <sub>17</sub> [TGA] <sub>1</sub> [TAGA] <sub>1</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
18	173 bp	262 bp	211 bp	[TAGA] <sub>18</sub> [TAGG][TG] <sub>1</sub>	Phillips et al. (2010)
18.3	176 bp	265 bp	214 bp	[TAGA] <sub>18</sub> [TGA] <sub>1</sub> [TAGA] <sub>1</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
19	177 bp	266 bp	215 bp	Not published	Asamura et al. (2008)
19.3	180 bp	269 bp	218 bp	[TAGA] <sub>19</sub> [TGA] <sub>1</sub> [TAGA] <sub>1</sub> [TAGG][TG] <sub>1</sub>	Lareu et al. (1998)
20.3	184 bp	273 bp	222 bp	Not published	Gamero et al. (2000)

15 NIST observed alleles circled in red  
from Appendix 1, J.M. Butler (2011) Advanced Topics in Forensic DNA Typing: Methodology

### NIST U.S. Population Allele Frequencies

#### D1S1656 (15 different alleles)

Allele	African American (N = 341)	Caucasian (N = 361)	Hispanic (N = 236)
10	0.01433	0.00277	0.00630
11	0.04871	0.07756	0.02731
12	0.06304	0.11773	0.08824
13	0.10029	0.06648	0.11555
14	<b>0.25788</b>	0.07895	0.11765
14.3	0.00716	0.00277	0.00420
15	0.15616	<b>0.14820</b>	0.13866
15.3	0.03009	0.05817	0.05042
16	0.11032	0.13573	<b>0.17437</b>
16.3	0.10029	0.06094	0.05462
17	0.02865	0.04709	0.04202
17.3	0.05014	0.13296	0.14496
18	0.00287	0.00554	0.00630
18.3	0.02436	0.05125	0.02521
19.3	0.00573	0.01385	0.00420

N = 938 (only unrelated samples used; fathers removed from this sample set) < 5/2N

### D1S1656 Characteristics

- 15 alleles observed
- 92 genotypes observed
- >89% heterozygotes (heterozygosity = 0.8934)
- 0.0220 Probability of Identity (P<sub>i</sub>)

$$P_i = \sum (genotype\ frequencies)^2$$

These values have been calculated for all 23 STR loci across the U.S. population samples examined

Loci sorted on Probability of Identity (P<sub>i</sub>) values

STR Locus	Alleles Observed	Genotypes Observed	Het. (obs)	P <sub>i</sub> value N = 938
SE33	53	292	0.9360	0.0069
Penta E*	20	114	0.8799	0.0177
D2S1338	13	68	0.8785	0.0219
D1S1656	15	92	0.8934	0.0220
D18S51	21	91	0.8689	0.0256
D12S391	23	110	0.8795	0.0257
FGA	26	93	0.8742	0.0299
Penta D*	16	71	0.8754	0.0356
D21S11	25	81	0.8358	0.0410
D19S433	16	76	0.8124	0.0561
D8S1179	11	45	0.7878	0.0582
vWA	11	38	0.8060	0.0622
D7S820	11	32	0.8070	0.0734
TH01	8	24	0.7580	0.0784
D16S539	9	28	0.7825	0.0784
D13S317	8	29	0.7655	0.0812
D10S1248	12	39	0.7825	0.0837
D2S441	14	41	0.7772	0.0855
D3S1358	11	30	0.7569	0.0873
D22S1045	11	42	0.7697	0.0933
CSF1PO	9	30	0.7537	0.1071
D5S818	9	34	0.7164	0.1192
TPOX	9	28	0.6983	0.1283

23 STR Loci present in STR kits rank ordered by their variability

Better for mixtures (more alleles seen)

There are several loci more polymorphic than the current CODIS 13 STRs

Better for kinship (low mutation rate)

### Random Match Probability for Various Combinations (assuming unrelated individuals)

STR Marker Combinations	RMP*	1 in ...
13 CODIS STRs	6.0E-16	1.7E+15
15 STRs (+D2S1338, D19S433)	7.3E-19	1.4E+18
18 STRs (+D2S441, D10S1248, D22S1045)	4.9E-22	2.0E+21
20 STRs (+D1S1656, D12S391)	2.8E-25	3.6E+24
23 STRs (+SE33, Penta D, Penta E)	1.2E-30	8.4E+29

\*RMP values calculated by combining Probability of Identity values for each locus

14 orders of magnitude improvement

### More Loci are Useful in Situations Involving Relatives

- Missing Persons and Disaster Victim Identification (kinship analysis)
- Immigration Testing (often limited references)
  - Recommendations for 25 STR loci
- Deficient Parentage Testing
  - often needed if only one parent and child are tested

Relationship testing labs are being pushed to answer more difficult genetic questions...and we want to make sure the right tools are in place

### In tomorrow's issue (February 25, 2011) of Forensic Science International...

Forensic Science International 205 (2011) 64-68

Contents lists available at ScienceDirect

Forensic Science International

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

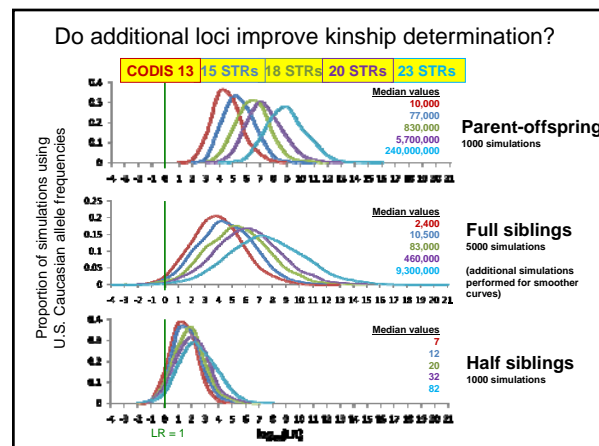
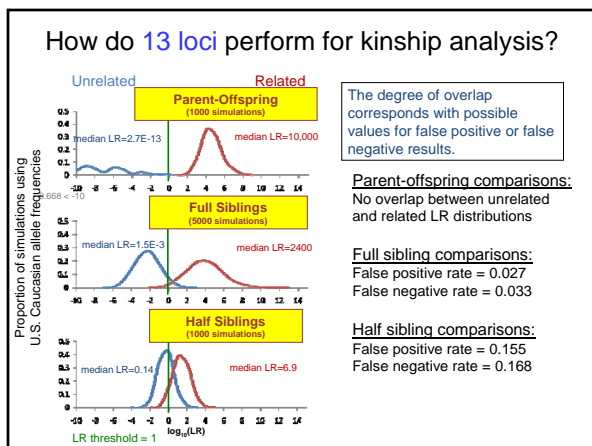
Examples of kinship analysis where Profiler Plus™ was not discriminatory enough for the identification of victims using DNA identification<sup>2</sup>

D. Hartman<sup>a,b,\*</sup>, L. Benton<sup>a</sup>, L. Morenos<sup>a</sup>, J. Beyer<sup>a</sup>, M. Spiden<sup>a</sup>, A. Stock<sup>a</sup>

<sup>a</sup>Victorian Institute of Forensic Medicine, 57-63 Kavanagh St. Southbank, VIC 3006, Australia  
<sup>b</sup>Department of Forensic Medicine, Monash University, 57-63 Kavanagh St. Southbank, VIC 3006, Australia

Disaster victim identification from the 2009 Victorian bushfires relied on DNA (82% involved kinship associations rather than direct matching)

They advocate additional autosomal STR loci to aid kinship associations



### Summary

- Additional autosomal STR loci exist in new STR kits and are being studied at NIST in U.S. population sample sets
- To avoid potential adventitious matches with large DNA databases, enable greater international data sharing, and aid missing persons applications, it is highly likely that additional loci will be added to the U.S. core in the future

### Thank you for your attention

#### Contact Information

**John Butler**  
NIST Fellow  
Group Leader of Applied Genetics  
[john.butler@nist.gov](mailto:john.butler@nist.gov)  
301-975-4049

<http://www.cstl.nist.gov/biotech/strbase>

**Our team publications and presentations are available at:**  
<http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm>

