Countdown to 2017: Internal Validation of the New CODIS Loci

Workshop: Countdown to 2017: Internal Validation of the New CODIS Loci
26th annual International Symposium on Human Identification
October 15, 2015
Grapevine, TX

NIST National Institute of Standards and Technology U.S. Department of Commerce

Agenda

1. Meet the new loci! (Becky Steffen)
2. Design of internal validation experiments (Peter Vallone)
3. Analysis and interpretation of validation data (Erica Romsos)
4. Examples of interpretation issues (Michael Coble)



Workshop Overview

- **Description:** With the adoption of new core STR loci, the validation and implementation of the new STR typing kits will be required by 2017. This workshop aims to review the new loci, discuss the design and analysis of experiments for internal validation, and illustrate interpretation issues. Examples of real validation data will be used throughout the workshop.
- **Brief review of new STR loci and commercially available STR typing kits**
- **Thoughts behind the design of internal validation experiments (the final application should guide the experiments)**
- **Examples of data analysis**
- **Discussion of specific interpretation issues**

Materials related to this workshop can be found on STRbase
http://www.cstl.nist.gov/strbase/training.htm

Disclaimer

We will mention commercial STR kit and instrument names, but we are in no way attempting to endorse any specific products.

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 it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Information presented does not necessarily represent the official position of the National Institute of Standards and Technology or the U.S. Department of Justice.

Acknowledgements

NIST
David Duewer
Margaret Kline
Steve Lund (NIST SED)
Sean Oliver (AFDIL)
Jo-Anne Bright (ESR)

Promega
Doug Storts

Contact Information
peter.vallone@nist.gov

Meet the New Loci:

An Introduction to the Additional Loci Beyond the 13 Core Loci

Becky Steffen
National Institute of Standards and Technology

Validation Workshop
Countdown to 2017: Internal Validation of the New CODIS Loci
Grapevine, TX
October 15, 2015

NIST National Institute of Standards and Technology U.S. Department of Commerce

Need for Additional STR Loci

- Larger DNA databases will require more loci
  - CODIS database currently has >14 million profiles and it continues to quickly grow
- To reduce the likelihood of adventitious matches in growing databases
- To increase international compatibility for data sharing efforts
- To increase discrimination power to aid missing persons cases

Expanded U.S. Core Loci


Forensic Science International: Genetics

Hares, D.R. (2012a) Expanding the CODIS core loci in the United States. *Forensic Sci. Int. Genet.* 6(1), e52-4.


Forensic Science International: Genetics

Hares, D.R. (2012b) Addendum to expanding the CODIS core loci in the United States. *Forensic Sci. Int. Genet.* 6(5), e135.

Proposed Required and Recommended CODIS Core Loci

Table 1
Revised ranked list of CODIS core loci.

Locus
Section A (required) **Required Loci**
Amelogenin
D18S51
FGA
D21S11
D8S1179
vWA
D13S317
D16S539
D7S820
TH01
D3S1358
D5S818
CSF1PO
D2S1338
D19S433
D1S1656
D12S391
D2S441
D10S1248
DYS391
Section B (in order of preference) **Recommended Loci**
TPOX
D22S1045
SE33

Penta D and Penta E were removed from this list in the addendum

Y-STR to confirm Amelogenin null alleles

No longer required

Consortium Validation Project (CVP)

October 2012 – July 2014

- The CVP involved a consortium of 11 CODIS laboratories representing casework, database and missing person laboratories
- Evaluated two STR multiplex kits that contain the proposed 20 CODIS core loci
  - Life Technologies' GlobalFiler® and GlobalFiler® Express PCR Amplification Kits
  - Promega PowerPlex® Fusion System
- Data from this validation project resulted in lessons learned and supported the final decision to expand the 13 CODIS core loci in the U.S. to 20

The Role of NIST in the CVP

- We provided 10 samples for the "Known Samples" portion of Component 1
- We worked closely with Doug Hares, the NDIS Custodian (oversees the U.S. National DNA Database), to evaluate all of the submitted data from each of the labs
- Dave Duewer has been instrumental in creating Excel macros to analyze the data and these programs are available for future validation studies
- We plan on co-publishing all of our findings in several manuscripts to the forensic community

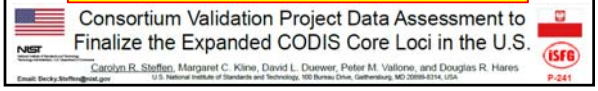
Validation Plan

	Consortium Validation Project								
	DATABASE			CASEWORK			MISSING PERSONS		
	# of Samples	Sample Type	Source	# of Samples	Sample Type	Source	# of Samples	Sample Type	Source
Component 1									
Spectral Confirmation									
Known Samples	10	Blood, buccal	NIST	10	blood/buccal	NIST	10	blood/buccal	NIST
Known or non-probative samples	45	Variable	In-House	25	blood, swabs, semen, bones, fluids, tissue	In-House	25	blood, swabs, bones, fluids, tissue, Trios (2 - 3 gen)	In-House
Precision	0	4,16,48 Allelic Ladder runs	In-House	0	4,16,48 Allelic Ladder runs	In-House	0	Allelic Ladder runs	In-House
Component 2									
Reproducibility	0	Will use known samples	In-house	0	Will use known samples	In-house	0	Will use known samples	In-house
Sensitivity and Stochastic	54	DNA extracts	In-House	54	DNA extracts	In House	54	DNA extracts	In-House
Mixture	16	DNA extracts	TBD In-house	96	DNA Extracts	TBD In-house	96	DNA Extracts	TBD In-house
Match Criteria	0	NA	NA	0	NA	NA	0	NA	NA
Contamination	0	NA	NA	0	NA	NA	0	NA	NA
Challenging Samples	25	Inhibited, degraded, Negative controls	In-house	20	Inhibited, degraded, Negative controls	In-House	20	Inhibited, degraded, Negative controls	In-House
Non-Probative Case Material	0	NA	NA	25	Buccal, Blood, Semen	In-house	25	bones, tissues	In-House
Component 3									
Population samples	50 or more	blood/buccal	In-House		blood/buccal	In-House		blood/buccal	In-House
Number of reactions	150			230			230		
Total number of amp pos/negatives (approx) =	20			20			20		
total number of reactions	170			250			250		

Validation Plan

- Some key areas of interest (performance of kits):
  - Genotype Concordance
  - Sensitivity
  - Precision
  - Non-Allelic PCR Products
  - Heterozygous Peak Balance
  - Stutter
  - Stochastic PCR Thresholds
  - Mixtures
  - Casework Challenge Samples

To see a summary of the results, see poster presentation from ISFG 2015: http://www.cstl.nist.gov/biotech/strbase/pub_pres/SteffenISFG2015CVPposter.pdf



Final Selection of the 20 Expanded CODIS Core Loci in the U.S.

Letter to the Editor

Selection and implementation of expanded CODIS core loci in the United States

System that contain the current 13 CODIS Core Loci, thus compelling the continued use of these loci in the core loci set. Review of the validation data did not result in exclusion of any of

20 CODIS core loci are required by January 1, 2017

Final selection of the 20 CODIS core loci was based on data driven decisions

Evolution of U.S. Core Loci

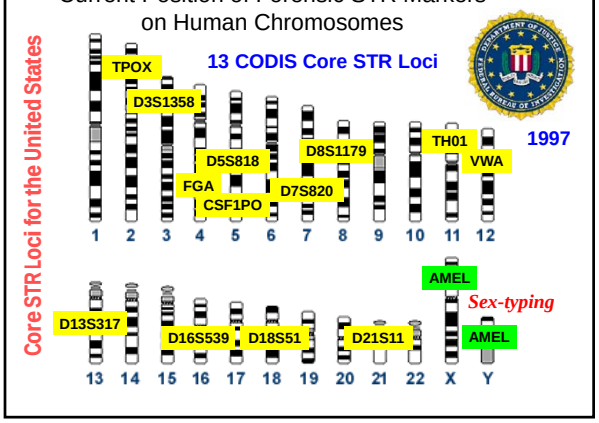
Required Loci (1997)	Proposed Required Loci (2012)	Required Loci (2015)
CSF1PO	CSF1PO	CSF1PO
D3S1358	D3S1358	D3S1358
D5S818	D5S818	D5S818
D7S820	D7S820	D7S820
D8S1179	D8S1179	D8S1179
D13S317	D13S317	D13S317
D16S539	D16S539	D16S539
D18S51	D18S51	D18S51
D21S11	D21S11	D21S11
FGA	FGA	FGA
TH01	TH01	TH01
TPOX	vWA	TPOX
vWA	D1S1656	vWA
Accepted Loci (later)	D2S441	D1S1656
D2S1338	D2S1338	D2S441
D19S433	D10S1248	D2S1338
Penta D	D12S391	D10S1248
Penta E	D19S433	D12S391
	DYS391	D19S433
	Amelogenin	D22S1045
	Recommended Loci (2012)	**Accepted Loci (2015)**
	TPOX	DYS391
	D22S1045	SE33
	SE33	Amelogenin
		Penta D
		Penta E

Original 13 Core U.S. STR Loci

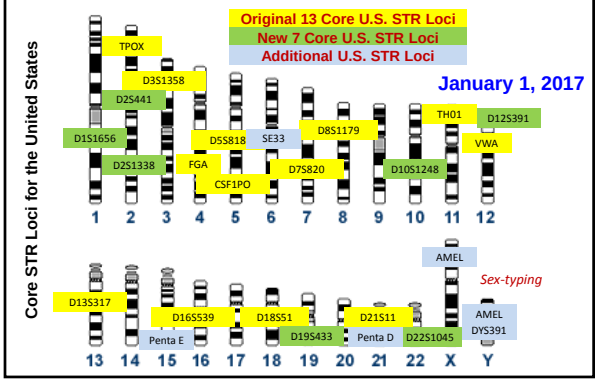
New Core U.S. STR Loci

Additional U.S. STR Loci

Current Position of Forensic STR Markers on Human Chromosomes



Future Position of U.S. Forensic Markers on Human Chromosomes



Conclusions from the CVP

- Review of the validation data did not result in exclusion of any of the current 13 CODIS Core Loci or the proposed additional core loci
- Based on the validation data, it was determined that the 20 loci that were in common between the two available PCR amplification kits would be selected as the new 20 CODIS Core Loci: the original 13 CODIS Core Loci, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 and D22S1045
- Amelogenin, DYS391, SE33, Penta D, and Penta E (also present in commercial STR kits) are accepted loci but are ***NOT*** required CODIS Core Loci

NIST U.S. Population Samples Used to Characterize the New Core Loci

NIST 1036 U.S. Population Samples

- 1032 males + 4 females
  - 361 Caucasians (2 female)
  - 342 African Americans (1 female)
  - 236 Hispanics
  - 97 Asians (1 female)

***Unrelated* samples**
All known or potential related individuals (based on autosomal & lineage marker testing) have been removed from the 1036 data set (e.g., only sons were used from father-son samples)

- Anonymous donors with self-identified ancestry
  - Interstate Blood Bank (Memphis, TN) – obtained in 2002
  - Millennium Biotech, Inc. (Ft. Lauderdale, FL) – obtained in 2001
  - DNA Diagnostics Center (Fairfield, OH) – obtained in 2007
- **Complete profiles with 23 autosomal STRs + PowerPlex Y23**
  - **Examined with multiple kits** and in-house primer sets enabling concordance
- Additional DNA results available on subsets of these samples
  - mtDNA control region/whole genome (AFDIL)
  - >100 SNPs (AIMs), 68 InDel markers, X-STRs (AFDIL)
  - NIST assays: miniSTRs, 26plex, >100 Y-STRs, 50 Y-SNPs

Data available on STRBase: **http://www.cstl.nist.gov/strbase/NISTpop.htm**

Benefits of NIST 1036 Data Set

- **Elimination of potential null alleles due to primer binding site mutations** through extensive concordance testing performed with different PCR primer sets from all available commercial STR kits
- **Ancestry testing performed** on DNA samples with autosomal SNPs, Y-SNPs, and mtDNA sequencing to verify self-declared ancestry categorization
- **Related individuals removed** based on Y-STR and mtDNA results

Characterization of STR Loci

13 CODIS + 10 Additional (23 total)

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‡	3 to 49	**178**

5 new European loci

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

25 Alleles Reported in the Literature for **D1S1656**

15 NIST observed alleles circled in red

Allele (Repeat #)	Promega ESX 17	Promega ESI 17	ABI NGM	Repeat Structure [TAGA]$_n$[TGA]$_{0-1}$[TAGA]$_n$TAGG[TG]$_5$	Reference
8	133 bp	222 bp	171 bp	[TAGA]$_8$[TG]$_5$	Phillips *et al.* (2010)
9	137 bp	226 bp	175 bp	[TAGA]$_9$[TG]$_5$	Phillips *et al.* (2010)
10 (a)	141 bp	230 bp	179 bp	[TAGA]$_{10}$[TG]$_5$	Lareu *et al.* (1998)
10 (b)	141 bp	230 bp	179 bp	[TAGA]$_{10}$TAGG[TG]$_5$	Phillips *et al.* (2010)
11	145 bp	234 bp	183 bp	[TAGA]$_{11}$[TG]$_5$	Lareu *et al.* (1998)
12 (a)	149 bp	238 bp	187 bp	[TAGA]$_{12}$[TG]$_5$	Lareu *et al.* (1998)
12 (b)	149 bp	238 bp	187 bp	[TAGA]$_{11}$TAGG[TG]$_5$	Lareu *et al.* (1998)
13 (a)	153 bp	242 bp	191 bp	[TAGA]$_{12}$TAGG[TG]$_5$	Lareu *et al.* (1998)
13 (b)	153 bp	242 bp	191 bp	[TAGA]$_{13}$[TG]$_5$	Phillips *et al.* (2010)
13.3	156 bp	245 bp	194 bp	[TAGA]$_3$TGA[TAGA]$_{11}$TAGG[TG]$_5$	Phillips *et al.* (2010)
14 (a)	157 bp	246 bp	195 bp	[TAGA]$_{13}$TAGG[TG]$_5$	Lareu *et al.* (1998)
14 (b)	157 bp	246 bp	195 bp	[TAGA]$_{14}$[TG]$_5$	Phillips *et al.* (2010)
14.3	160 bp	249 bp	198 bp	[TAGA]$_4$TGA[TAGA]$_9$TAGG[TG]$_5$	Phillips *et al.* (2010)
15	161 bp	250 bp	199 bp	[TAGA]$_{14}$TAGG[TG]$_5$	Lareu *et al.* (1998)
15.3	164 bp	253 bp	202 bp	[TAGA]$_4$TGA[TAGA]$_{10}$TAGG[TG]$_5$	Lareu *et al.* (1998)
16	165 bp	254 bp	203 bp	[TAGA]$_{15}$TAGG[TG]$_5$	Lareu *et al.* (1998)
16.3	168 bp	257 bp	206 bp	[TAGA]$_4$TGA[TAGA]$_{11}$TAGG[TG]$_5$	Lareu *et al.* (1998)
17	169 bp	258 bp	207 bp	[TAGA]$_{16}$TAGG[TG]$_5$	Lareu *et al.* (1998)
17.1	170 bp	259 bp	208 bp	Not published	Schröer *et al.* (2000)
17.3	172 bp	261 bp	210 bp	[TAGA]$_4$TGA[TAGA]$_{12}$TAGG[TG]$_5$	Lareu *et al.* (1998)
18	173 bp	262 bp	211 bp	[TAGA]$_{17}$TAGG[TG]$_5$	Phillips *et al.* (2010)
18.3	176 bp	265 bp	214 bp	[TAGA]$_4$TGA[TAGA]$_{13}$TAGG[TG]$_5$	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]$_4$TGA[TAGA]$_{14}$TAGG[TG]$_5$	Lareu *et al.* (1998)
20.3	184 bp	273 bp	222 bp	Not published	Gamero *et al.* (2000)

from Appendix 1, J.M. Butler (2011) Advanced Topics in Forensic DNA Typing: Methodology

NIST **U.S. Population Allele Frequencies**

D1S1656 (15 different alleles)

15 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
10	0.0146	0.0000	0.0028	0.0064
11	0.0453	0.0309	0.0776	0.0275
12	0.0643	0.0464	0.1163	0.0890
13	0.1009	0.1340	0.0665	0.1144
14	**0.2573**	0.0619	0.0789	0.1165
14.3	0.0073	0.0000	0.0028	0.0042
15	0.1579	**0.2784**	**0.1496**	0.1377
15.3	0.0292	0.0000	0.0582	0.0508
16	0.1096	0.2010	0.1357	**0.1758**
16.3	0.1023	0.0155	0.0609	0.0508
17	0.0278	0.0722	0.0471	0.0424
17.3	0.0497	0.0876	0.1330	0.1483
18	0.0029	0.0155	0.0055	0.0064
18.3	0.0234	0.0515	0.0499	0.0254
19.3	0.0073	0.0052	0.0152	0.0042

N=1036
(only unrelated samples used; fathers removed from this sample set)

D1S1656 Characteristics

- **15 alleles** observed
- **93 genotypes** observed
- **>89% heterozygotes** (heterozygosity = 0.8890)
- **0.0224 Probability of Identity** (P_I)

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

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

Loci sorted on Probability of Identity (P_I) values

Locus	Alleles Observed	Genotypes Observed	Het (obs)	P_I Value n=1036
SE33	52	304	0.9353	0.0066
Penta E	23	138	0.8996	0.0147
D2S1338	13	68	0.8793	0.0220
D1S1656	15	93	0.8890	0.0224
D18S51	22	93	0.8687	0.0258
D12S391	24	113	0.8813	0.0271
FGA	27	96	0.8745	0.0308
Penta D	16	74	0.8552	0.0382
D21S11	27	86	0.8330	0.0403
D8S1179	11	46	0.7992	0.0558
D19S433	16	78	0.8118	0.0559
vWA	11	39	0.8060	0.0611
D7S820	11	32	0.7944	0.0726
D16S539	9	28	0.7761	0.0749
D13S317	8	29	0.7674	0.0765
TH01	8	24	0.7471	0.0766
D2S441	15	43	0.7828	0.0841
D10S1248	12	39	0.7819	0.0845
D3S1358	11	30	0.7519	0.0915
D22S1045	11	44	0.7606	0.0921
CSF1PO	9	31	0.7558	0.1054
D5S818	9	34	0.7297	0.1104
TPOX	9	28	0.6902	0.1358

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

N=1036
(only unrelated samples used)

There are several **additional loci** that are more polymorphic than the **original CODIS 13 STR loci**

361 Caucasians
342 African Americans
236 Hispanics
97 Asians

Probability of Identity Combinations
(assuming unrelated individuals)

STR Kit or Core Set of Loci	Number of Loci	Total PI* N=1036	Magnitude Improvement
CODIS 13	**13**	5.02E-16	n/a
ESS 12	**12**	3.04E-16	n/a
CODIS 20	**20**	9.35E-24	10^8
GlobalFiler	23	7.73E-28	10^{12}
Investigator 24plex QS	23	7.73E-28	10^{12}
PowerPlex Fusion	24	6.58E-29	10^{13}
PowerPlex Fusion 6C	27	1.80 x 10-32	10^{17}

*PI values do not include Amelogenin or Y-indel

Meet the New Loci

- General information
  - Chromosomal location
  - Repeat motif
  - Allele range and number of alleles

Reported in the literature and summarized in Appendix 1, J.M. Butler (2011) Advanced Topics in Forensic DNA Typing: Methodology

- Population statistics using NIST 1036 data set
  - Observed alleles, genotypes and heterozygosity
  - Probability of Identity values
  - Allele frequencies

Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U.S. population data for 29 autosomal STR loci. Forensic Sci. Int. Genet. 7: e82-e83.

- STR Multiplex Kits
  - Marker position: size and color

Meet the New Loci… D1S1656

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
D1S1656	1q42	TAGA	8 to 20.3	25

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
15	93
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.8890	0.0224

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	SID	Purple	154-209
PowerPlex Fusion	Fluorescein	Blue	151-207
PowerPlex Fusion 6C	FL-6C	Blue	152-209
Investigator 24plex QS	BTG	Green	155-201

U.S. Population Allele Frequencies

15 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
10	0.0146	0.0000	0.0028	0.0064
11	0.0453	0.0309	0.0776	0.0275
12	0.0643	0.0464	0.1163	0.0890
13	0.1009	0.1340	0.0665	0.1144
14	**0.2573**	0.0619	0.0789	0.1165
14.3	0.0073	0.0000	0.0028	0.0042
15	0.1579	**0.2784**	**0.1496**	0.1377
15.3	0.0292	0.0000	0.0582	0.0508
16	0.1096	0.2010	0.1357	**0.1758**
16.3	0.1023	0.0155	0.0609	0.0508
17	0.0278	0.0722	0.0471	0.0424
17.3	0.0497	0.0876	0.1330	0.1483
18	0.0029	0.0155	0.0055	0.0064
18.3	0.0234	0.0515	0.0499	0.0254
19.3	0.0073	0.0052	0.0152	0.0042

Meet the New Loci… D2S441

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
D2S441	2p14	TCTA/TCAA	8 to 17	22

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
15	43
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.7828	0.0841

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	NED	Yellow	75-113
PowerPlex Fusion	Fluorescein	Blue	208-247
PowerPlex Fusion 6C	FL-6C	Blue	211-252
Investigator 24plex QS	BTR	Red	79-133

U.S. Population Allele Frequencies

15 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
8.0	0.0000	0.0052	0.0000	0.0000
9.0	0.0029	0.0000	0.0014	0.0000
9.1	0.0000	0.0052	0.0014	0.0000
10.0	0.0848	0.2680	0.2105	**0.3369**
11.0	**0.3626**	**0.3505**	**0.3435**	0.2987
11.3	0.0439	0.0309	0.0609	0.0445
12.0	0.1652	0.2165	0.0471	0.0360
12.3	0.0058	0.0052	0.0042	0.0021
13.0	0.0439	0.0258	0.0291	0.0233
13.3	0.0029	0.0000	0.0000	0.0000
14.0	0.2675	0.0825	0.2410	0.2055
14.3	0.0015	0.0000	0.0000	0.0000
15.0	0.0190	0.0103	0.0596	0.0487
16.0	0.0000	0.0000	0.0014	0.0021
17.0	0.0000	0.0000	0.0000	0.0021

Meet the New Loci… D2S1338

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
D2S1338	2q35	TGCC/TTCC	10 to 31	40

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
13	68
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.8793	0.0220

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	SID	Purple	275-356
PowerPlex Fusion	JOE	Green	218-299
PowerPlex Fusion 6C	JOE-6C	Green	221-304
Investigator 24plex QS	BTY	Yellow	360-438

U.S. Population Allele Frequencies

13 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
15.0	0.0015	0.0000	0.0014	0.0000
16.0	0.0556	0.0206	0.0374	0.0297
17.0	0.1009	0.0567	**0.1856**	0.1695
18.0	0.0424	0.1340	0.0734	0.0805
19.0	**0.1389**	**0.1804**	0.1205	**0.1928**
20.0	0.1038	0.1598	0.1565	0.1271
21.0	0.1360	0.0155	0.0374	0.0318
22.0	0.1374	0.0515	0.0346	0.0572
23.0	0.1038	0.1649	0.1053	0.1398
24.0	0.0833	0.1289	0.1150	0.0763
25.0	0.0775	0.0464	0.1025	0.0784
26.0	0.0146	0.0361	0.0305	0.0169
27.0	0.0044	0.0052	0.0000	0.0000

Meet the New Loci… D10S1248

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
D10S1248	10q26.3	GGAA	7 to 19	13

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
12	39
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.7819	0.0845

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	SID	Purple	80-132
PowerPlex Fusion	Fluorescein	Blue	248-295
PowerPlex Fusion 6C	FL-6C	Blue	254-302
Investigator 24plex QS	BTY	Yellow	83-143

U.S. Population Allele Frequencies

12 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
8.0	0.0029	0.0000	0.0000	0.0000
9.0	0.0029	0.0000	0.0000	0.0000
10.0	0.0073	0.0052	0.0000	0.0000
11.0	0.0351	0.0000	0.0014	0.0042
12.0	0.1301	0.0876	0.0319	0.0424
13.0	0.2339	**0.3196**	**0.3075**	0.2733
14.0	**0.2763**	0.2526	0.2978	**0.3390**
15.0	0.1974	0.2062	0.1967	0.2119
16.0	0.0877	0.0979	0.1330	0.0996
17.0	0.0249	0.0309	0.0277	0.0254
18.0	0.0015	0.0000	0.0014	0.0021
19.0	0.0000	0.0000	0.0028	0.0021

Meet the New Loci… D12S391

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
D12S391	12p13.2	AGAT/AGAC	13 to 27.2	52

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
24	113
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.8813	0.0271

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	SID	Purple	211-270
PowerPlex Fusion	CXR-ET	Red	131-191
PowerPlex Fusion 6C	CXR-6C	Red	130-190
Investigator 24plex QS	BTG	Green	203-265

U.S. Population Allele Frequencies

24 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
14.0	0.0015	0.0000	0.0000	0.0000
15.0	0.0775	0.0412	0.0319	0.0445
16.0	0.0673	0.0103	0.0222	0.0424
17.0	0.1667	0.0825	0.1274	0.0763
17.1	0.0044	0.0000	0.0000	0.0000
17.3	0.0044	0.0000	0.0208	0.0169
18.0	**0.2529**	**0.2629**	**0.1717**	0.1780
18.1	0.0015	0.0000	0.0000	0.0000
18.3	0.0044	0.0000	0.0249	0.0127
19.0	0.1477	0.1753	0.1247	**0.1886**
19.1	0.0088	0.0000	0.0000	0.0021
19.3	0.0044	0.0052	0.0042	0.0064
20.0	0.1038	0.1959	0.1108	0.1547
20.1	0.0029	0.0000	0.0000	0.0000
20.3	0.0000	0.0000	0.0000	0.0021
21.0	0.0643	0.0979	0.1288	0.1123
22.0	0.0365	0.0567	0.0956	0.0678
22.2	0.0000	0.0000	0.0000	0.0021
23.0	0.0292	0.0258	0.0693	0.0572
24.0	0.0132	0.0103	0.0471	0.0169
24.3	0.0000	0.0052	0.0000	0.0000
25.0	0.0088	0.0155	0.0166	0.0064
26.0	0.0000	0.0103	0.0028	0.0064
27.0	0.0000	0.0052	0.0014	0.0064

Meet the New Loci… D19S433

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
D19S433	19q12	AAGG/TAGG	5.2 to 20	36

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
16	78
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.8118	0.0559

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	NED	Yellow	115-173
PowerPlex Fusion	CXR-ET	Red	191-256
PowerPlex Fusion 6C	CXR-6C	Red	192-255
Investigator 24plex QS	BTY	Yellow	213-279

U.S. Population Allele Frequencies

16 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
9.0	0.0000	0.0000	0.0000	0.0021
10.0	0.0102	0.0000	0.0014	0.0021
11.0	0.0629	0.0000	0.0055	0.0148
12.0	0.1228	0.0361	0.0706	0.0657
12.2	0.0365	0.0258	0.0014	0.0127
13.0	**0.2456**	0.2835	0.2548	0.2225
13.2	0.0526	0.0206	0.0069	0.0445
14.0	0.2105	**0.2990**	**0.3615**	**0.3538**
14.2	0.0746	0.1031	0.0235	0.0381
15.0	0.0804	0.0619	0.1565	0.1356
15.2	0.0614	0.1237	0.0360	0.0551
16.0	0.0044	0.0103	0.0568	0.0254
16.2	0.0263	0.0309	0.0152	0.0275
17.0	0.0000	0.0000	0.0069	0.0000
17.2	0.0088	0.0052	0.0014	0.0000
18.2	0.0029	0.0000	0.0014	0.0000

Meet the New Loci... D22S1045

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
D22S1045	22q12.3	ATT	7 to 20	14

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
11	44
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.7606	0.0921

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	TAZ	Red	83-127
PowerPlex Fusion	CXR-ET	Red	420-472
PowerPlex Fusion 6C	CXR-6C	Red	430-478
Investigator 24plex QS	BTY	Yellow	144-191

U.S. Population Allele Frequencies

11 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
8.0	0.0073	0.0000	0.0000	0.0000
10.0	0.0409	0.0000	0.0000	0.0148
11.0	0.1447	0.2010	0.1399	0.0636
12.0	0.0541	0.0052	0.0125	0.0127
13.0	0.0029	0.0000	0.0069	0.0085
14.0	0.0775	0.0103	0.0568	0.0275
15.0	**0.2515**	**0.3093**	**0.3213**	**0.4258**
16.0	0.1915	0.2268	0.3823	0.3496
17.0	0.2091	0.2216	0.0748	0.0911
18.0	0.0146	0.0206	0.0055	0.0064
19.0	0.0058	0.0052	0.0000	0.0000

Meet the New Loci... Penta D

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
Penta D	21q22.3	AAAGA	1.1 to 19	50

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
16	74
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.8552	0.0382

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
PowerPlex Fusion	JOE	Green	370-461
PowerPlex Fusion 6C	JOE-6C	Green	373-470

U.S. Population Allele Frequencies

16 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
2.2	0.1140	0.0000	0.0042	0.0169
3.2	0.0088	0.0000	0.0000	0.0021
5.0	0.0439	0.0000	0.0000	0.0064
6.0	0.0102	0.0103	0.0042	0.0021
7.0	0.0439	0.0000	0.0042	0.0021
8.0	0.1082	0.0052	0.0208	0.0191
9.0	0.1681	**0.3144**	0.2216	**0.2415**
10.0	0.0994	0.1856	0.1150	0.1568
11.0	**0.1798**	0.1804	0.1260	0.1568
12.0	0.1082	0.1804	**0.2327**	0.1631
13.0	0.0833	0.1031	0.1967	0.1441
13.4	0.0015	0.0000	0.0000	0.0000
14.0	0.0249	0.0103	0.0609	0.0720
15.0	0.0044	0.0103	0.0097	0.0106
16.0	0.0000	0.0000	0.0028	0.0042
17.0	0.0015	0.0000	0.0014	0.0021

Meet the New Loci... Penta E

STR Locus	Location	Repeat Motif	Allele Range	# Alleles
Penta E	15q26.2	AAAGA	5 to 32	53

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
23	138
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.8890	0.0224

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
PowerPlex Fusion	Fluorescein	Blue	354-475
PowerPlex Fusion 6C	FL-6C	Blue	362-482

U.S. Population Allele Frequencies

23 different alleles

Allele	African American (n=342)	Asian (n=97)	Caucasian (n=361)	Hispanic (n=236)
5.0	0.0950	0.0722	0.0762	0.0360
6.0	0.0015	0.0000	0.0000	0.0000
7.0	0.1038	0.0000	0.1690	0.1186
8.0	**0.1667**	0.0052	0.0139	0.0254
9.0	0.0512	0.0258	0.0125	0.0169
10.0	0.0468	0.0619	0.0859	0.0847
11.0	0.0643	0.1598	0.0873	0.0742
12.0	0.1287	0.0928	**0.1994**	**0.1737**
13.0	0.1038	0.0515	0.0859	0.0932
14.0	0.0687	0.0515	0.0623	0.0720
15.0	0.0556	**0.1134**	0.0429	0.0911
15.4	0.0000	0.0000	0.0014	0.0021
16.0	0.0409	0.0567	0.0512	0.0614
17.0	0.0439	0.0825	0.0485	0.0551
18.0	0.0161	0.0773	0.0332	0.0339
19.0	0.0073	0.0412	0.0152	0.0212
19.4	0.0000	0.0103	0.0000	0.0000
20.0	0.0044	0.0206	0.0097	0.0212
21.0	0.0000	0.0206	0.0028	0.0064
22.0	0.0000	0.0309	0.0014	0.0021
23.0	0.0000	0.0206	0.0000	0.0064
24.0	0.0000	0.0000	0.0014	0.0000
25.0	0.0015	0.0052	0.0000	0.0042

Meet the New Loci… SE33

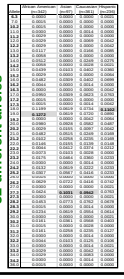
STR Locus	Location	Repeat Motif	Allele Range	# Alleles
SE33	6q14	AAAG*	3 to 49	178

*SE33 alleles have a complex repeat structure

Population Statistics using NIST 1036 data set

Alleles Observed	Genotypes Observed
52	304
Heterozygosity Observed	Probability of Identity (P_I) Value n=1036
0.9353	0.0066

52 different alleles



U.S. Population Allele Frequencies

Locations in common U.S. STR kits

STR Multiplex Kit	Dye Label	Color	Size (bp)
GlobalFiler	TAZ	Red	306-444
PowerPlex Fusion 6C	CXR-6C	Red	270-429
Investigator 24plex QS	BTG	Green	272-434

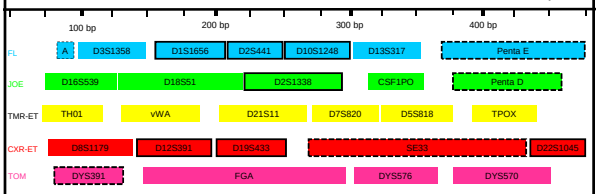
NIST U.S. Population Data

- The data from our 1036 U.S. population samples is currently available on STRBase:
  http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm
- A summary of the NIST 1036 data set has been published in Profiles in DNA for autosomal and YSTR loci
- Population Data announcements have been published in FSI: Genetics for
  - 29 autosomal STR loci (*Hill et al*)
  - 23 Y-STR loci (*Coble et al*)

PowerPlex Fusion 6C – the largest commercial STR multiplex kit available for CE use

Promega PowerPlex **FUSION 6C**

27plex



- 27 STR loci in 6 dyes (3500 use or 3130 upgrade required)
  - Includes SE33 and 3 Y-STR markers (DYS391, DYS570, DYS576)
  - Includes Penta D and E
- Direct amplification and casework capabilities: 60 min amp for both (one kit)
- PowerPlex Fusion gives ~17 orders of magnitude improvement using the NIST 1036 data set

http://www.promega.com/products/pm/genetic-identity/powerplex-fusion/

SRM 2391c: PCR-Based DNA Profiling Standard


National Institute of Standards & Technology
Certificate of Analysis
Standard Reference Material® 2391c
PCR-Based DNA Profiling Standard

- Contains the only samples (Components A-F) **certified** for all 27 of the loci contained in PowerPlex Fusion 6C, including autosomal and Y-STR loci
- Certified genotypes/haplotypes have been assigned by electrophoretic base pair (bp) size matched to sequenced alleles and also confirmed by direct Sanger sequencing

SRM 2391c can be used as "known" samples for validation studies of new kits

Summary

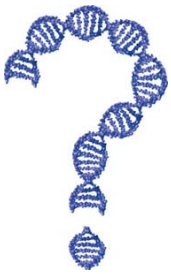
- Additional STR loci are important as DNA databases grow larger each year: the power of discrimination increases as new loci are added
  - Adding seven new loci (CODIS 13 vs CODIS 20) adds approximately 8 orders of magnitude improvement
- NIST has a set of 1036 unrelated U.S. population samples that have been used to fully characterize 23 autosomal STR loci available in commercial STR multiplex kits
- Commercial companies are continuing to release larger STR multiplex kits to meet the needs of the forensic community and these must be validated before use

Acknowledgments

Dr. David Duewer

Dr. Doug Hares

Contact info:
becky.steffen@nist.gov
301-975-4275



Design of internal validation experiments

Peter M. Vallone, Ph.D.
Leader, Applied Genetics Group

Workshop: Countdown to 2017: Internal Validation of the New CODIS Loci
26th annual International Symposium on Human Identification
October 15, 2015
Grapevine, TX

NIST National Institute of Standards and Technology
U.S. Department of Commerce

Background

- Take you through our process of developing and carrying out an internal validation plan for PowerPlex Fusion 6C
  - Why Fusion 6C – new for us

- Context: *at NIST we are primarily typing single source DNA extracts for population and concordance studies*
- We do not run casework samples or undergo audits
- Not meant to be a recipe – improvements can be made

Team of People

- Pete, Mike, Becky, Erica

- Steve Lund

  - (NIST Statistical Engineering Division)

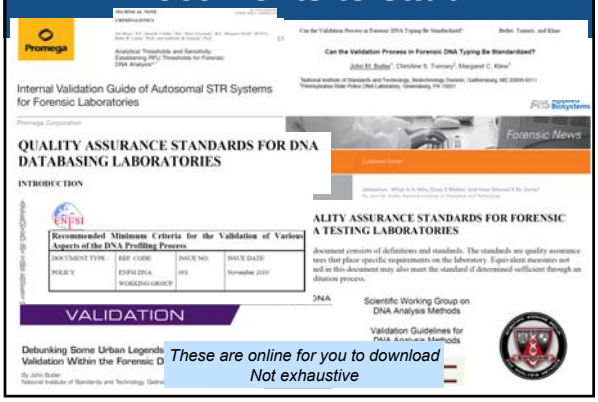
Dave Duewer (NIST data analysis)

- Sean Oliver AFDIL
- Jo-Anne Bright E/S/R THE SCIENCE BEHIND THE TRUTH

Getting started – Homework

- Refresh our knowledge of internal validation
- Previous validation articles, talks, and workshops
  - What is new?
- Draft internal validation experiments
  - Discuss and revise

Documents to start



These are online for you to download
Not exhaustive

Initial Parameter Testing

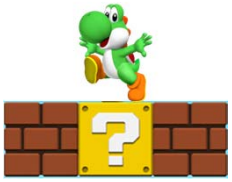
- 'Test drive' the new STR typing kit
- Determine # of PCR cycles, injection time/voltage, and optimal DNA input amount
- This lays the groundwork for the validation experiments
- Carefully coordinate and quantitate candidate samples to be used in the validation work
- Running on a 24 capillary 3500xL; Veriti thermal cycler

DNA Template Quantitation

- The importance of selecting and characterizing the validation samples
- Use the same quantitation kit as one you have already validated
- For the sensitivity and mixture studies to be meaningful the samples must be well characterized (LOD, ratios)
- Our candidate validation samples were quantitated with PowerQuant and assigned a value against the SRM 2372 A calibration standard

Top 10 unanswerable questions

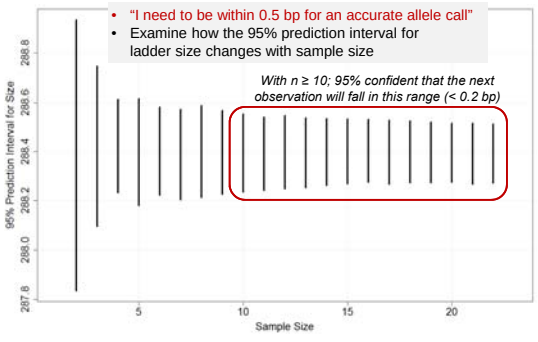
1. **How many samples do I to need to run for my validation?**
2. What is the meaning of life?
3. Did Tony Soprano die?
4. Is there a God?
5. Do blondes have more fun?
6. What is the best way to lose weight?
7. Is there anybody out there?
8. How many licks does it take to get to the center of a lollipop?
9. What is the secret to happiness?
10. How long will I live?



Confidence

Discrete measurements e.g. base pair sizing

- "I need to be within 0.5 bp for an accurate allele call"
- Examine how the 95% prediction interval for ladder size changes with sample size



Confidence
Kit/assay concordance

- Testing results from Kit A to Kit B – do I get the same genotype?
  - Different: primers; mastermix; cycling conditions
  - Typing multiple samples will provide confidence in the overall the genotyping process
    - Test different population groups, number of alleles, separation between alleles (in bp), challenging separations, etc
  - Compare profiles and assess concordance across kits
  - Balancing cost and time

e.g. < 99.3%

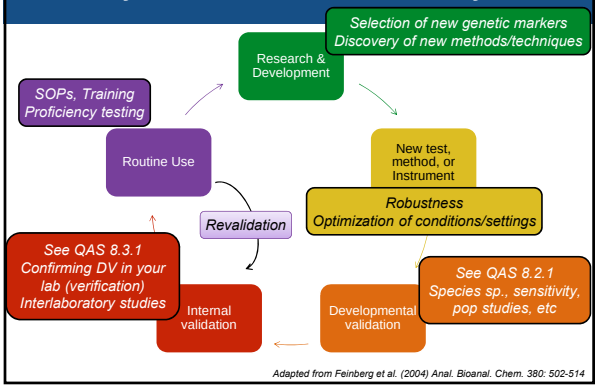
Confidence
Genotyping

Steve Lund

- **Rule of 3:** When a particular outcome is not observed in the first n independent "trials," the observer can be 95% confident that the long-term (i.e. after MANY trials) proportion of future trials in which the particular outcome is observed is ***less than 3/n***.
- Quick illustrations:
- If you watch me make 100 free throws in a row, it's reasonable for you to conclude that I'm not worse than a 97% free-throw shooter (3/100 = .03).
- If you observed no discordances in 10,000 allele designations, it's reasonable to conclude the rate of concordance is at least 99.97%.

This may be a reasonable means to assess confidence in genotyping
Samples*loci*alleles = n
50*27*2 = 2,700; 3/2,700 = 99.89%...is this useful? Maybe?

Lifecycle of a method of analysis

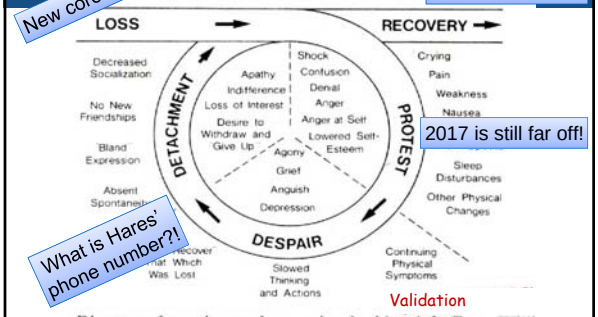


Adapted from Feinberg et al. (2004) Anal. Bioanal. Chem. 380: 502-514

Validation

New core loci?!

New kit is online!



2017 is still far off!

What is Hares' phone number?!

. Diagram of reaction cycle associated with ~~grief.~~ (From Willi Lamers, M.D. Reproduced by permission.)[1]

Internal Validation

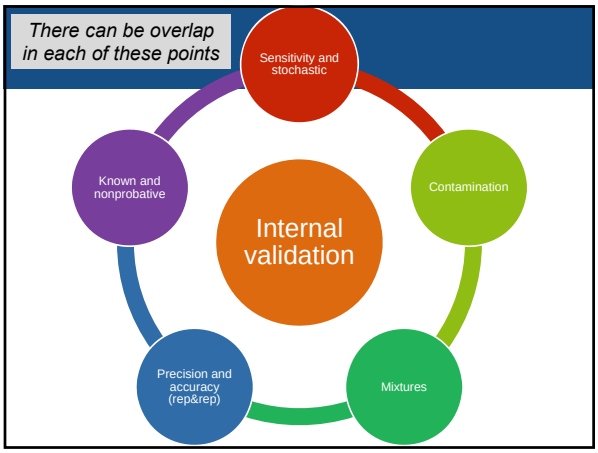
- The internal validation process shall include the studies detailed below. If conducted within the same laboratory, developmental validation studies may satisfy some elements of the internal validation guidelines. The laboratory should evaluate the appropriate sample number and type, based on the methodology and/or application necessary to demonstrate the potential limitations and reliability. The laboratory should determine the suitability of each study based on the methodology and may determine that a study is not necessary.

SWGDAM validation guidelines 2012

Internal Validation

1. **Known and nonprobative evidence samples or mock evidence samples**
2. **Sensitivity and Stochastic Studies**
3. **Precision and accuracy** **(Repeatability/Reproducibility)**
4. **Mixture studies**
5. **Contamination assessment**

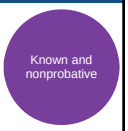
SWGDAM validation guidelines 2012



Known and nonprobative evidence samples or mock evidence samples

- Methods intended for casework samples should be evaluated and tested using known samples and nonprobative evidence samples or mock case samples. Methods intended for database samples should be evaluated and tested using known samples. Results from these studies should be compared to the previous results of known samples and/or nonprobative evidence or mock case samples to **ensure concordance**.

Known and nonprobative evidence samples or mock evidence samples



- Why?
  - Genotype concordance
  - Baseline stutter and peak height balance values
  - Will the technique work with my 'type' of sample?
- Comments
  - Nonprobative or mock evidence: **NIST population samples**
  - Knowns: SRM 2391c can be useful (esp. with the newer STR loci)
  - Commercial controls, ATCC, Coriell, blood – **master set concept**
  - How many should be run?

Sensitivity and Stochastic Studies

- The laboratory should demonstrate sensitivity levels of the test. Sensitivity studies are used to determine the dynamic range, ideal target range, limit of detection, limit of quantitation, heterozygote balance (e.g., peak height ratio) and the signal to noise ratio associated with the assay. Sensitivity studies can also be used to evaluate excessive random (stochastic) effects generally resulting from low quantity and/or low quality samples.

Sensitivity and Stochastic Studies

- Why?
  - Dynamic range of STR kit
  - Optimal DNA input target range
  - Analytical threshold
  - Stochastic threshold (and effects/artifacts)
- Comments
  - Select samples with varying alleles and degrees of heterozygosity
  - Well characterized samples quantitated with appropriate qPCR kit
  - Dilution range?

Precision and accuracy of the assay should be demonstrated

- **Precision** characterizes the degree of mutual agreement among a series of individual measurements, values and/or results. Precision depends only on the distribution of random errors and does not relate to the true value or specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results.
- **Accuracy** is the degree of conformity of a measured quantity to its actual (true) value. Accuracy of a measuring instrument is the ability of a measuring instrument to give responses close to a true value.

Repeatability: Precision and accuracy of results (e.g., quantitative and/or qualitative) of the same operator and/or detection instrument should be evaluated.
Reproducibility: Precision and accuracy of results (e.g., quantitative and/or qualitative) among different operators and/or detection instruments should be evaluated.

Precision and Accuracy
Reproducibility and Repeatability

Precision and accuracy (rep&rep)

- Why
  - **Precision**: for STR typing we seem to mean bp sizing
  - Important for sizing peaks and assigning alleles
  - **Accuracy**: 'true value' – similar to genotype concordance
- Comments
  - Allelic ladders (Precision)
  - Positive control, knowns (Accuracy)
  - Repeatability: same sample typed multiple times (use the positive control)
  - Reproducibility: same samples run by different operators (nonprobative)

Mixture Studies

- Mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated. These studies will assist a casework laboratory to establish guidelines for mixture interpretation, which may include determination of the number of contributors to the mixture, determination of the major and minor contributor profiles, and contributor ratios or proportions. **A simplified mixture study may also assist a databasing laboratory to recognize mixtures and/or contamination**.

Mixture studies

Mixtures

- Why
  - General LOD for minor contributors in a mixture
  - Should mirror contamination LOD
  - Development of mixture interpretation guidelines
- Comments
  - Samples with varying degrees of allele overlap
  - Again, samples must be well characterized and quantitated
  - Varying ratios and contributors. Set total target DNA template amount
  - Basic study – more needed for mixture interpretation validation?

Contamination assessment

- The laboratory should evaluate, using both controls and known samples, the detection of exogenous DNA (including allele drop-in and heteroplasmy) originating from reagents, consumables, operator and/or laboratory environment.

Contamination assessment

- Why
  - **Reagents**
  - Consumables
  - Operator
  - Laboratory environment
- Comments
  - Negative controls (TE/water)
  - Not reagent blanks, not CE checkerboards, zebra stripes
  - Focusing on the STR kit reagents; not CE crosstalk

Internal validation plan

After discussing these Internal Validation topics…

- Sat down and discussed the experimental plan for validation, appropriate sample types, number of samples, dilutions, mixture ratios, and replicates
- Reference to previous studies, developmental validation studies, discussions with other labs performing validations
- Iterative process within our group

Master List of Samples

- NIST population samples (35)
- SRM 2391c (six samples) *44 total*
- 9947A, 9948, 2800M (three samples)

- Allelic ladder from kit

Known and nonprobative evidence samples or mock evidence samples

44 unique samples

Known and nonprobative

- 35 NIST population samples
- 9947A and 9948
- Positive control 2800M
- 6 samples from SRM 2391c (knowns)

Known and nonprobative evidence samples or mock evidence samples

1 injection on 3500xL
1 ladder
1 pos/neg control

Known and nonprobative

	1	2	3
A	SRM 2391c A	Ladder	
B	SRM 2391c B		
C	SRM 2391c C		
D	SRM 2391c D		
E	SRM 2391c E		
F	SRM 2391c F		
G	+		
H	-		

Known and nonprobative evidence samples or mock evidence samples

2 injections on 3500xL
2 ladders
2 pos/neg controls

Known and nonprobative

	1	2	3	4	5	6
A	Ladder	9	17	25	33	1
B	2	10	18	26	34	16
C	3	11	19	27	35	24
D	4	12	20	28	9947a (36)	
E	5	13	21	29	9948 (37)	
F	6	14	22	30	+	
G	7	15	23	31	-	
H	8	+	-	32	Ladder	

Operator One

Positive control = 2800M

Known and nonprobative evidence samples or mock evidence samples

2 injections on 3500xL
2 ladders
2 pos/neg controls

Known and nonprobative

7	8	9	10	11	12
Ladder	9	17	25	33	1
2	10	18	26	34	16
3	11	19	27	35	24
4	12	20	28	9947a (36)	
5	13	21	29	9948 (37)	
6	14	22	30	+	
7	15	23	31	-	
8	+	-	32	Ladder	

Operator Two

Precision and accuracy (rep&rep)

Performed (amp and injection) by a different operator
Reproducibility

Sensitivity and Stochastic Studies

Sensitivity and stochastic

- 3 unique samples
  - High levels of heterozygosity
  - Subset of our nonprobatives (sample # 29, 31, 32)
- Run in triplicate
  - Three unique amplifications of the serial dilutions
- Dilution points
  - 2000, 1000, 500, 250, 125, 62.5, 31.3 pg

Sensitivity and Stochastic Studies

3 injections on 3500xL
3 ladders
3 pos/neg control

Sensitivity and stochastic

	1	2	3	4	5	6	7	8	9
A	2.00	+	2.00	-	2.00	Ladder	2.00	Ladder	2.00
B	1.00	0.03	1.00	0.03	1.00	0.03	1.00	0.03	1.00
C	0.50	0.06	0.50	0.06	0.50	0.06	0.50	0.06	0.50
D	0.25	0.13	0.25	0.13	0.25	0.13	0.25	0.13	0.25
E	0.13	0.25	0.13	0.25	0.13	0.25	0.13	0.25	0.13
F	0.06	0.50	0.06	0.50	0.06	0.50	0.06	0.50	0.06
G	0.03	1.00	0.03	1.00	0.03	1.00	0.03	1.00	0.03
H	Ladder	2.00	-	2.00	+	2.00	-	2.00	+
	Sample 29			Sample 31			Sample 32		

Sensitivity and Stochastic Studies

3 injections on 3500xL
3 ladders
3

Sensitivity and stochastic

Comments on sensitivity (from RSO)

- For low template work test down to 15 pg
- For databasing labs test higher limits to confirm upper end (for 'blown out' signal)
- Sensitivity is a function of STR kit/lab/CE parameters (can't just rely on Dev. Validation)
- It is important for the internal validation to assess this performance

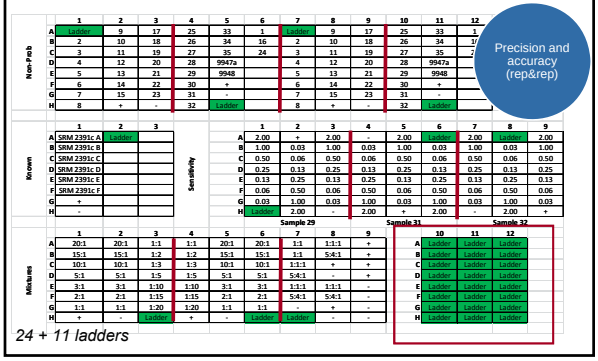
	9
A	2.00
B	1.00
C	0.50
D	0.25
E	0.13
F	0.06
G	0.03
H	+

Precision and Accuracy
Reproducibility and Repeatability

Precision and accuracy (rep&rep)

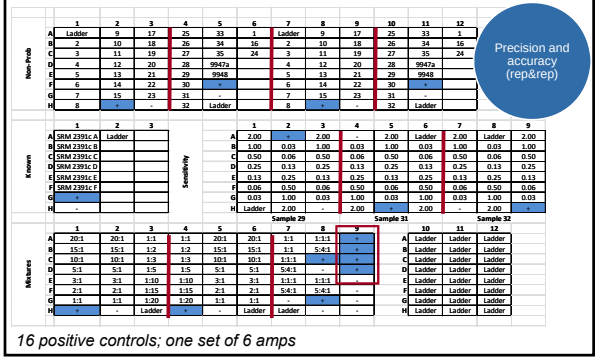
- 35 allelic ladders were run
  - One ladder per injection (11)
  - One injection of just ladders (24) (to cover all capillaries)
- 6 + 37 knowns/nonprobative (genotype accuracy)
- 16 Positive controls
  - A positive was typed per every 24 samples (one per injection) for repeatability
  - Repeatability: positive also typed six times in a single experiment
  - 37 samples were typed by another operator as part of the reproducibility study – as discussed in the known and nonprobative section

Precision and Accuracy
Reproducibility and repeatability



24 + 11 ladders

Precision and Accuracy
Reproducibility and repeatability



16 positive controls; one set of 6 amps

Mixture studies

- Three sets of two person mixtures (male)
- One set of 9947A and 9948 1:1 (example)
- Two examples of three person mixtures
- Target 1 ng total per PCR
- Ratios
  - Two person: 20, 15, 10, 5, 3, 2, 1 to 1; amplified in duplicate
  - Three person: 1:1:1 and 5:4:1; amplified in duplicate

Mixtures

Mixture studies

2 injections on 3500xL
2 ladders
2 pos/neg control
Three, two person mixtures

Mixtures

	1	2	3	4	5	6
A	20:1	20:1	1:1	1:1	20:1	20:1
B	15:1	15:1	1:2	1:2	15:1	15:1
C	10:1	10:1	1:3	1:3	10:1	10:1
D	5:1	5:1	1:5	1:5	5:1	5:1
E	3:1	3:1	1:10	1:10	3:1	3:1
F	2:1	2:1	1:15	1:15	2:1	2:1
G	1:1	1:1	1:20	1:20	1:1	1:1
H	+	-	Ladder	+	-	Ladder
	Samples 27:28		**Samples 35:33**		**Samples 8:2**	

Mixture studies

1 injection on 3500xL
1 ladder
6 pos/7 neg control

Mixtures

9947A:9948

Samples 9:2:7

Samples 22:16:6

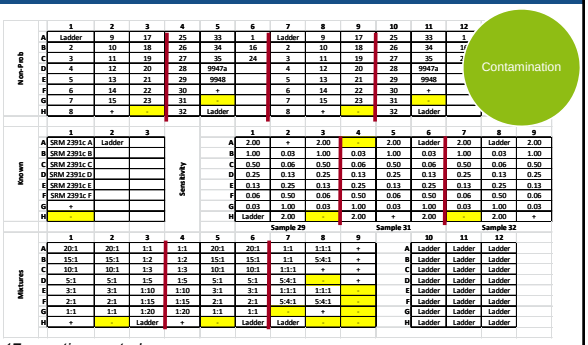
7	8	9
1:1	1:1:1	+
1:1	5:4:1	+
1:1:1	+	+
5:4:1	-	+
1:1:1	1:1:1	-
5:4:1	5:4:1	-
-	+	-
Ladder	-	-

Contamination assessment

Contamination

- 17 negative controls (TE) were interspersed throughout the study
- Focusing on the STR reagent kits
- Evaluate for allele and elevated noise

Contamination assessment



17 negative controls

Summary Table

	Unique Samples	Amps	Approx ng of DNA
Nonprob	37	74	2 per sample
Known	6	6	1 per sample
Positive	1	16	1 per sample
Sensitivity	0 (subset)	63	12 x3
Mixtures	0 (subset)	52	15 x4
Negative	0	17	0
Ladders	1	0	0
	45	228	< 200 ng

12 total injections on the 3500xL

Approximate Times

- Discussion meetings
  - 5 hours (spread over 2-3 meetings)
- Selection and characterization of validation samples
  - 3 days
- Initial parameter testing experiments
  - 5 days (including data analysis)
- Carrying out the validation experiments
  - 1 week
- Data analysis
  - 2 weeks (GeneMapper analysis, exporting, parsing data, excel)

Comments

- Discussions between members of the validation team were invaluable
  - Pull in new people as needed
  - Asking Why? versus just How many?
- Create a sizeable pool of well characterized samples
  - This will not be your last internal validation
- An opportunity for members of the lab to have buy in on the experiments that will become part of SOP

Acknowledgements

NIST
David Duewer
Margaret Kline
Steve Lund (NIST SED)
Sean Oliver (AFDIL)
Jo-Anne Bright (ESR)

Promega
Doug Storts

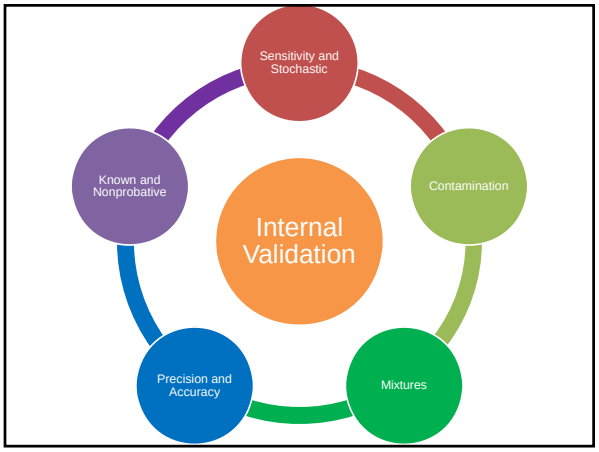
Contact Information
peter.vallone@nist.gov

Analysis of Validation Data

Erica Romsos
National Institute of Standards and Technology

Validation Workshop
Countdown to 2017: Internal Validation of the New CODIS Loci
Grapevine, TX
October 15, 2015

NIST National Institute of Standards and Technology U.S. Department of Commerce



Assumptions Being Made

- All samples have been extremely well quantified prior to testing
  - PowerQuant chemistry
  - All candidate samples tested in triplicate
  - SRM 2372 Component A used for producing the standard curve (in triplicate)
    - Precise standard curve produced for all samples (within the limits expected for slope, intercept and R^2)

Initial Parameter Testing

- 7 samples were selected for initial testing
  - Diluted to 0.5 ng/µL and 1.0 ng/µL
- Amplified in triplicate using PowerPlex Fusion 6C
  - 28 cycles, 29 cycles, 30 cycles
- Injected on 3500xL at 1.2 kV
  - 24 seconds, 15 seconds, 8 seconds

Initial Parameter Testing: Signal

Per the manufacturer's guidelines the signal range targeted was 4000-8000 RFU

DNA Input Amount and Cycle Number	Injection Time 8s	15s	24s	Average Observed RFU Values
0.5ng_28cyc	613	1473	2433	
0.5ng_29cyc	1123	2309	3695	
0.5ng_30cyc	2592	4795	7900	
1.0ng_28cyc	1080	2338	4231	
1.0ng_29cyc	3015	5773	9886	
1.0ng_30cyc	4950	8699	13642	

Amplification and detection instrumentation may vary. You may need to optimize protocols including the amount of template DNA, cycle number, injection conditions and loading volume for your laboratory instrumentation. Testing at Promega shows that 29 cycles works well for 1.0ng of purified DNA templates.

Initial Parameter Testing: Peak Height Ratios

DNA Input Amount	Injection Time 8s	15s	24s	Average Observed Peak Height Ratio
0.5ng_28cyc	89.0%	88.9%	89.0%	
0.5ng_29cyc	89.6%	89.5%	89.5%	
0.5ng_30cyc	86.8%	86.8%	86.7%	
1.0ng_28cyc	92.4%	92.4%	92.3%	
1.0ng_29cyc	91.7%	91.8%	91.8%	
1.0ng_30cyc	90.9%	90.9%	90.8%	

Peak height ratios were less than 90% with 0.5 ng DNA input at all cycling and injection parameters

Initial Parameter Testing: Decision Making

- Comparison of signal strength at each parameter and peak height ratios
- Signal was within the "sweet spot" at 1.0 ng DNA input, 29 cycles, and 15 second injection at 1.2 kV
- Highest peak height ratios were observed at 1.0 ng DNA input, 28 cycles, and 15 second injection at 1.2 kV
  - Signal was low at these parameters (~2300 RFU)

Final Parameters

Signal Strength

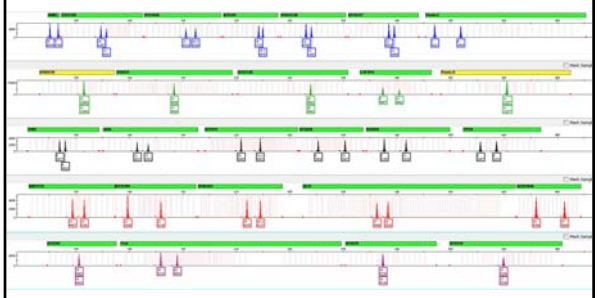
DNA Input Amount and Cycle Number	Injection Time 8s	15s	24s	Average Observed RFU Values
0.5ng_28cyc	613	1473	2433	
0.5ng_29cyc	1123	2309	3695	
0.5ng_30cyc	2592	4795	7900	
1.0ng_28cyc	1080	2338	4231	
1.0ng_29cyc	3015	5773	9886	
1.0ng_30cyc	4950	8699	13642	

Peak Height Ratios

DNA Input Amount	Injection Time 8s	15s	24s	Avg Obs PHR
1.0ng_28cyc	92.35%	92.45%	92.34%	
1.0ng_29cyc	91.72%	91.82%	91.80%	
1.0ng_30cyc	90.91%	90.90%	90.83%	

1.0 ng DNA input, 29 cycles, and 15 second injection at 1.2 kV was selected for moving forward with internal validation testing

Representative Electropherogram



1.0 ng input DNA
15 s injection @ 1.2 kV

From here we move into our internal validation strategy

Slide provided by Robin Cotton

The data is here!

Plan A

Plan B

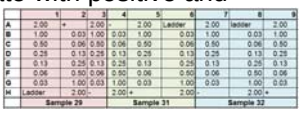
Slide provided by Robin Cotton

So, o...

Sensitivity Testing: Experimental Design

Sensitivity and Stochastic

- Three mostly heterozygous samples selected
  - 112 individual alleles represented
- Diluted from stock to the following DNA input amounts:
  - 2.0 ng, 1.0 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0.0625 ng, and 0.031 ng
- Amplified in triplicate with positive and negative controls

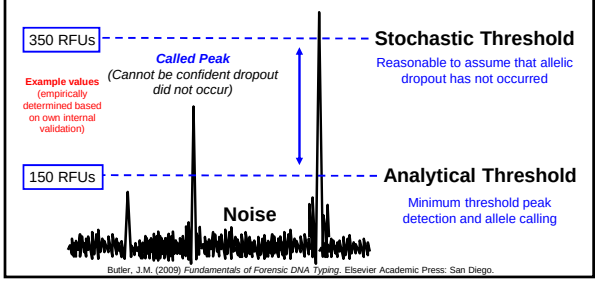


	1	2	3	4	5	6	7	8	9
A	2.00	+	2.00	-	2.00	Ladder	2.00	ladder	2.00
B	1.00	0.03	1.00	0.03	1.00	0.03	1.00	0.03	1.00
C	0.50	0.06	0.50	0.06	0.50	0.06	0.50	0.06	0.50
D	0.25	0.13	0.25	0.13	0.25	0.13	0.25	0.13	0.25
E	0.13	0.25	0.13	0.25	0.13	0.25	0.13	0.25	0.13
F	0.06	0.50	0.06	0.50	0.06	0.50	0.06	0.50	0.06
G	0.03	1.00	0.03	1.00	0.03	1.00	0.03	1.00	0.03
H	Ladder	2.00 -		2.00 +		2.00 -		2.00 +	
	Sample 29			Sample 31			Sample 32		

Sensitivity Testing: Goals

Sensitivity and Stochastic

- Establishment of thresholds
  - Analytical and stochastic thresholds

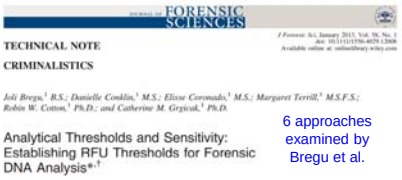


Butler, J.M. (2009) *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press: San Diego.

Slide provided by Jo-Anne Bright @ ESR

Setting Thresholds

Sensitivity and Stochastic


JOURNAL OF FORENSIC SCIENCES

TECHNICAL NOTE

CRIMINALISTICS

Joli Bregu,[1] B.S.; Danielle Conklin,[1] M.S.; Elisse Coronado,[1] M.S.; Margaret Terrill,[1] M.S.F.S.; Robin W. Cotton,[1] Ph.D.; and Catherine M. Grgicak,[1] Ph.D.

Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis*,†

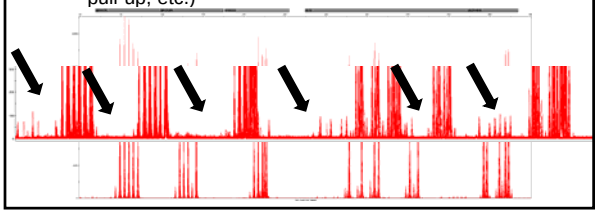
6 approaches examined by Bregu et al.

- Bregu et al. "Results showed that when a substantial mass of DNA (i.e., >1 ng) was amplified, the baseline noise increased, suggesting the **application of an AT derived from negatives should only be applied to samples with low levels of DNA**"
- "…it is **recommended that color specific thresholds be utilized**"

Analytical Threshold: Methodology

Sensitivity and Stochastic

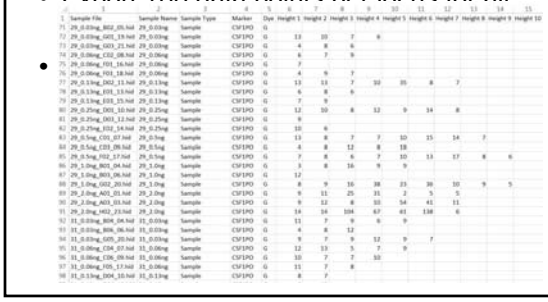
- Baseline noise values calculated from sensitivity dilution series
  - Threshold set at 1 RFU for all dye channels
  - Remove calls for all alleles and artifacts (stutter, n+4, pull-up, etc.)



Analytical Threshold: Methodology

Sensitivity and Stochastic

- Export 100 data points per locus for all
-



Analytical Threshold: Methodology

Sensitivity and Stochastic

- Calculated the average noise per dye channel

=AVERAGE('Raw Data Export (For Dye)'!R2C6:R484C105)

Select all data for each dye set and calculate average, standard deviation, minimum, and maximum within Excel

Analytical Threshold: Calculations

Sensitivity and Stochastic

AT = Average Noise + 10(SD)

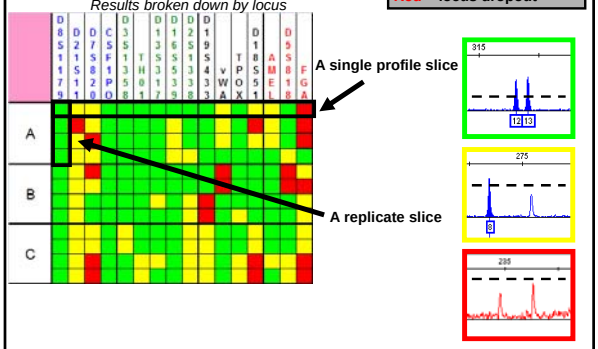
Dye Specific Analytical Thresholds

	Average	Stdev	Min	Max
FL	11	11.5	1	104
JOE	14	13.9	2	241
TMR-ET	11	11.0	1	160
CXR-ET	13	15.2	2	154
TOM	9	6.4	1	88

Heat Map Explanation

Green = full (correct) type
Yellow = allele dropout
Red = locus dropout

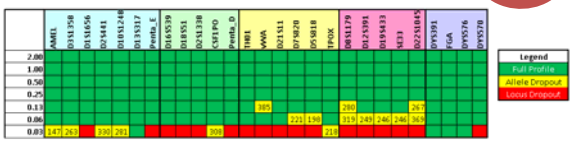
Results broken down by locus



This is an easy way to look at a lot of data at once

Sensitivity Testing: Combined Heat Map

Sensitivity and Stochastic



Data was analyzed with a sensitivity macro program developed by Dave Duewer, but can be done by hand within Excel by recording all allelic and full locus dropout

Allelic Dropout first observed at 125 pg of DNA input in at least one replicate of each of the three individual sample

Full locus dropout first observed at 62 pg of DNA input in at least one replicate of each of the three individual samples

Stochastic Threshold: Methodology

Sensitivity and Stochastic

- Analyzed data from the sensitivity study with dye specific analytical thresholds
- Examined sample amounts where dropout was observed (125 pg, 62 pg, and 31 pg)
  - Used to examine stochastic effects including severe imbalance of heterozygous alleles and allele dropout

70 heterozygous loci were evaluated

Stochastic Threshold: Calculations

Sensitivity and Stochastic

Stochastic Threshold: The RFU value of highest surviving false homozygous peak per dye channel

	Blue	Green	Yellow	Red	Purple
min	135	159	125	167	79
max	452	431	385	369	357
average	209	240	166	232	184
stdev	71	81	52	53	128

Summary of Thresholds

PowerPlex Fusion 6C: 15 sec @ 1.2 kV (29 cycles; 1.0 ng DNA Input)			
	AT (RFU)	Highest Surviving Peak (RFU)	ST (RFU)
FL	130	452	455
JOE	155	431	435
TMR-ET	125	385	385
CXR-ET	165	369	370
TOM	75	357	360

Thresholds need to be reestablished when changes to DNA input, injection time, injection voltage, and number of cycles are altered

Peak Height Ratios

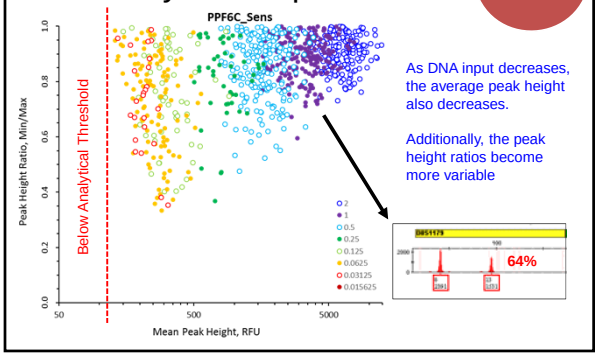
- All samples at 1.0 ng of DNA input included in PHR calculation
  - Positive controls (n=16)
  - Non-probative (n=37)
  - Known (n=5)
  - 1.0 ng sensitivity (n=9)

Median peak height ratio above 87% across all 1.0 ng DNA samples (n=67)

Locus	Median
TPOX	0.939
D5S818	0.934
D13S317	0.927
D16S539	0.922
D1S1656	0.917
D18S51	0.913
D2S441	0.910
TH01	0.909
D10S1248	0.909
D8S1179	0.908
D7S820	0.908
vWA	0.902
D19S433	0.901
D21S11	0.901
AMEL	0.901
Penta_E	0.899
Penta_D	0.898
SE33	0.898
D2S1338	0.896
D22S1045	0.892
D3S1358	0.892
FGA	0.889
D12S391	0.884
CSF1PO	0.871

Peak Height Ratios: By DNA Input Amount

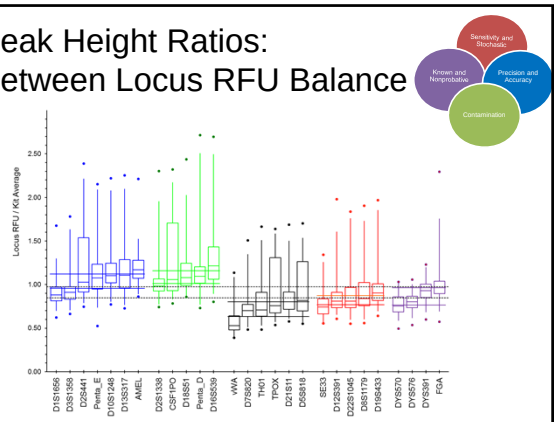
Sensitivity and Stochastic



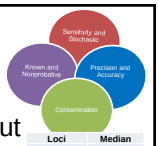
As DNA input decreases, the average peak height also decreases.

Additionally, the peak height ratios become more variable

Peak Height Ratios: Between Locus RFU Balance



Stutter Percentages

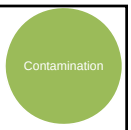


- All samples at 1.0 ng of DNA input included in PHR calculation
  - Positive controls (n=16)
  - Non-probative (n=37)
  - Known (n=5)
  - 1.0 ng sensitivity (n=9)

Stutter percentages ranged from 1.88% (Penta D) to 15.55% (SE33)

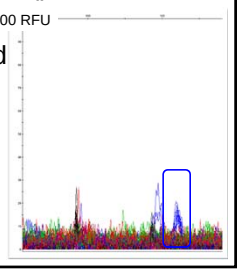
Loci	Median
Penta_D	1.88
TH01	2.01
Penta_E	3.38
TPOX	3.59
D2S441	5.21
D13S317	5.69
D7S820	6.43
D5S818	6.68
CSF1PO	6.81
DYS391	6.89
D19S433	7.10
D16S539	7.34
D8S1179	7.56
FGA	7.85
D21S11	8.16
D2S1338	8.34
vWA	8.38
DYS570	8.53
D10S1248	8.60
D1S1656	8.62
D3S1358	8.84
DYS576	8.90
D18S51	8.92
D22S1045	8.99
D12S391	9.01
SE33	15.55

Contamination



- 17 negative controls evaluated
- No observed contamination (peaks above AT)
- Pull-up from ILS observed

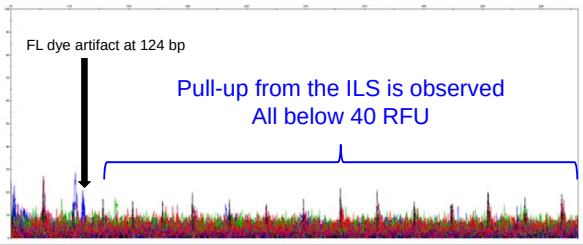
Artifact observed ~124 bp in the FL dye in all negative controls



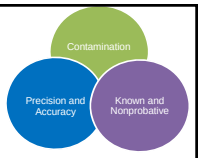
Negative Controls

Contamination

Overlay of 17 negative controls
RFU scale is set at 100 RFU



Concordance



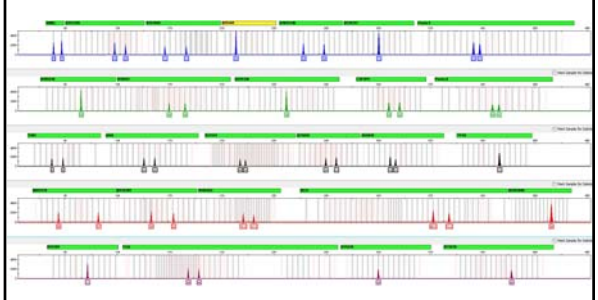
- 37 non-probative samples
  - Amplified and separated twice
- 6 known samples
- 16 positive controls
  - One positive control run 16 times

100 % concordance across 44 individual samples

250 unique alleles typed

SRM 2391c-Component C

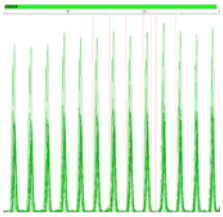
Known and Nonprobative



Precision and Accuracy

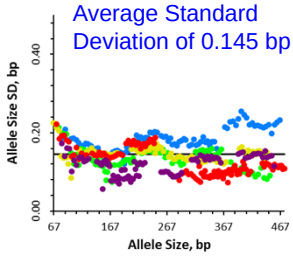
Precision and Accuracy

- 35 allelic ladders analyzed
  - 27 loci
  - 17,325 alleles



Overlay of 35 allelic ladders

Average Standard Deviation of 0.145 bp

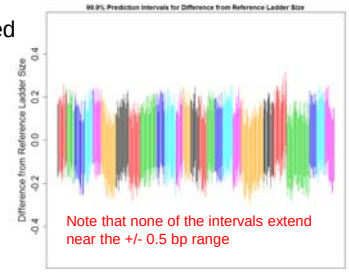


Precision and Accuracy

Precision and Accuracy

One injection of 24 ladders performed

- 1 ladder assigned as the "ladder"
- 22 ladders assigned as samples
- 1 ladder failed



Note that none of the intervals extend near the +/- 0.5 bp range

How wide of an interval would you need to capture 99.9% of future allele calls?

Repeatability

Precision and Accuracy | Known and Nonprobative

- Positive control typed in every injection
- One injection with 6 positive controls

- Total of 16 positive controls typed by a single operator

100 % concordance across 16 positive controls

Reproducibility

Precision and Accuracy | Known and Nonprobative

- 37 non-probative samples amplified and separated independently
  - Two different operators

	1	2	3	4	5	6	7	8	9	10	11	12
a	ladder	9	17	25	33	1	ladder	9	17	25	33	1
b	2	10	18	26	34	16	2	10	18	26	34	16
c	3	11	19	27	35	24	3	11	19	27	35	24
d	4	12	20	28	9947a		4	12	20	28	9947a	
e	5	13	21	29	9948		5	13	21	29	9948	
f	6	14	22	30	+		6	14	22	30	+	
g	7	15	23	31	-		7	15	23	31	-	
h	8	+	-	32	ladder		8	+	-	32	ladder	
	Person 1						Person 2					

100 % concordance between both operators and across all samples

Mixtures

Mixtures

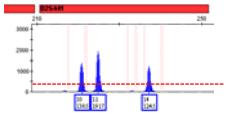
- 1.0 ng of total DNA input for all mixtures
- Three 2-person mixtures created
  - Low allele sharing, moderate allele sharing, high allele sharing
- Two 3-person mixtures created
  - Low allele sharing and high allele sharing
  - 1:1:1 and 5:4:1

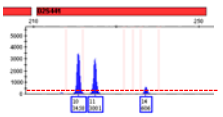
1:1
1:2
1:3
1:5
1:10
1:15
1:20

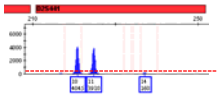
2 Person Mixtures

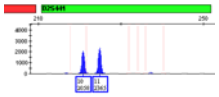
High allele sharing

Mixtures









Stochastic Threshold for the FL dye is 455 RFU

Minor component falls below the stochastic threshold at 1:15 and drops out completely at 1:20

2 Person Mixtures

Moderate allele sharing

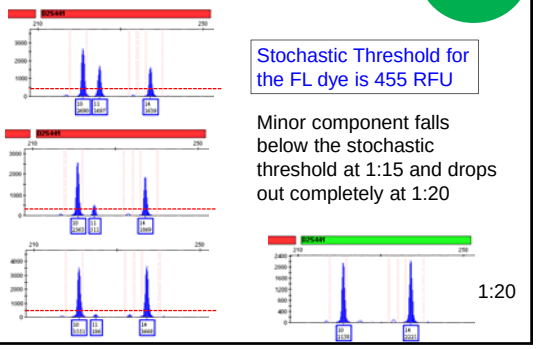
Mixtures

1:1

1:5

1:15

1:20



Stochastic Threshold for the FL dye is 455 RFU

Minor component falls below the stochastic threshold at 1:15 and drops out completely at 1:20

2 Person Mixtures

Low allele sharing

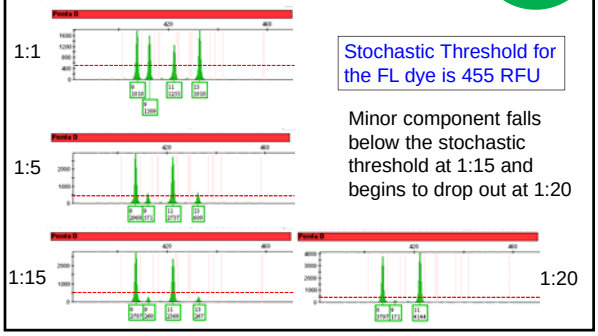
Mixtures

1:1

1:5

1:15

1:20



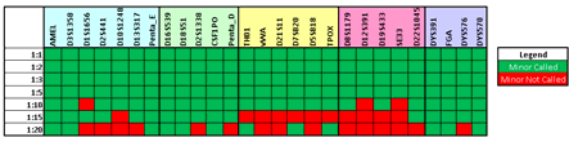
Stochastic Threshold for the FL dye is 455 RFU

Minor component falls below the stochastic threshold at 1:15 and begins to drop out at 1:20

2 Person Mixtures

Heat map for calling minor contributor

Mixtures



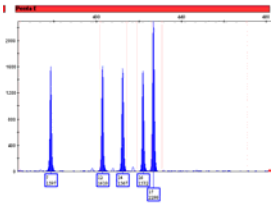
Dropout of the minor component was first observed within the 1:10 low allele sharing mixture at three loci

Dropout of the minor component was consistently observed in the 1:15 and 1:20 mixtures

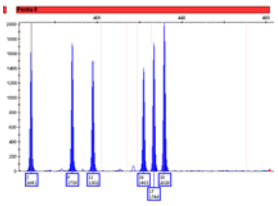
3 Person Mixtures

Mixtures

Mixture A



Mixture B



More than a 2 person mixture...

Mike will discuss interpretation further….

NIST Software Tools Used

Developed by Dave Duewer (NIST)

- In-house software programs have been developed to assist with data analysis and concordance studies
  - Peak Height Ratios, stutter, allele frequencies, population statistics, concordance
- Recently, additional programs have been developed to aid in the FBI Consortium Validation Project (CVP)
  - Precision, sensitivity and stochastic, mixture, challenging and non-probative samples
- **These programs can be utilized during validation studies**

http://www.cstl.nist.gov/strbase/pub_pres/ValidationWebinar-Hill-Aug2014.pdf

New NIST Software Tools

Developed by Dave Duewer (NIST)

From NIST STRBase Website:

- **STR_AlleleFreq**
  - Allele frequencies
  - Peak height ratios
  - Inter-locus balance
- **STR_StutterFreq**
  - Stutter frequencies
  - Locus-specific stutter
  - Allele-specific stutter
- **STR_Genotype**
  - Population statistics
  - Allele frequencies

http://www.cstl.nist.gov/biotech/strbase/software.htm

Multiple tools within Excel used to evaluate data

General Summary

1.0 ng DNA input, 29 cycles, and 15 second injection at 1.2 kV

	AT (RFU)	Highest Surviving Peak (RFU)	ST (RFU)
FL	130	452	455
JOE	155	431	435
TMR-ET	125	385	385
CXR-ET	165	369	370
TOM	75	357	360

Peak height ratios above 70% for 98% of the data at 1.0ng DNA input

Stutter percentages ranged from 2% to 15%

Precision within 0.15 base pairs

100% Concordance (for knowns, positives, and nonprobatives)

Minor component within mixtures can be detected at 1:15

Conclusions

- Validation is a necessary evil within any laboratory
  - With the implementation of the new CODIS loci, everyone is doing it (so you're not alone)
- The data generated from your internal validation determine **interpretation guidelines** and sets the groundwork for writing your SOPs
- With any changes to your initial protocols (cycle number, injection time, injection voltage, etc) you need to revalidate to determine appropriate thresholds for data analysis and interpretation
  - The new STR kits are more sensitive and may require additional testing at the lower bounds

Acknowledgments

Dr. David Duewer (NIST)
Dr. Charlotte Word (Consultant)
Dr. Robin Cotton (BU)
Dr. Jo Bright (ESR)

Contact info:
erica.romsos@nist.gov
+1-301-975-5107

Interpretation Issues

Michael Coble
National Institute of Standards and Technology

Validation Workshop
Countdown to 2017: Internal Validation of the New CODIS Loci

ISHI INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION GRAPEVINE, TX • OCT 12-15, 2015

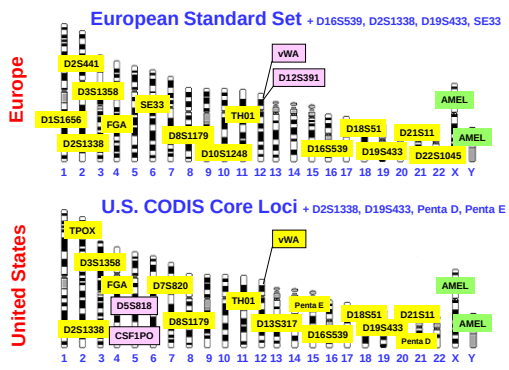
Grapevine, TX
October 15, 2015

NIST National Institute of Standards and Technology U.S. Department of Commerce

Some things to know…

- Syntenic STRs
- One-base separation issues
- Stutter issues

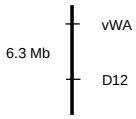
Common Forensic STR Loci



Syntenic Loci – vWA and D12S391

Table 2
Inter-STR physical separation, genetic distances in cM and Kosambi-adjusted recombination fractions (Rc) for the 20 closest STR pairs.

Chrom.	STR pair	dbSNP identifiers or repeat region/ neighbour SNPs	SNP position	Physical distance in nucleotides	cM interval of closest HapMap SNP proxies	Rc from Kosambi mapping function
6	SE33	rs71021371	88,986,927	3,462,987	4.4141	0.0440
	D6S1043	rs111544865	92,449,914			
21	D21S2055	rs113225349	41,191,444	3,865,246	9.9114	0.0978
	Penta D	rs7279663	45,056,212			
12	vWA	rs10579907	6,093,104	6,357,030	11.9414	0.1172
	D12S391	rs113002069	12,450,134			
8	D8S1132	rs71307053	107,328,946	18,583,159	16.4798	0.1591
	D8S1179	rs67563232	125,907,114			
6	D6S1043	rs111544865	92,449,914	20,430,123	18.7992	0.1796
	D6S474	rs113991233	112,879,130			



Chr 12

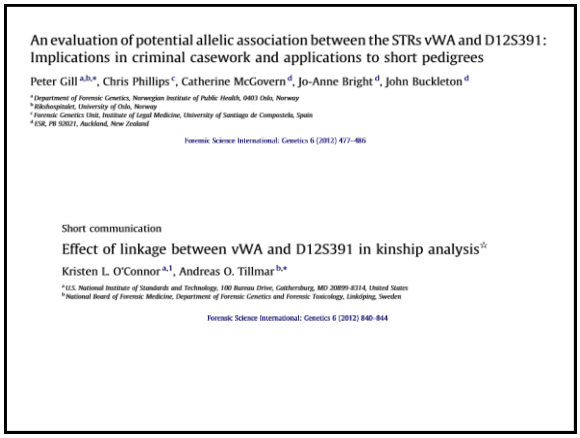
Phillips et al. FSI:G (2012)

Syntenic Loci – vWA and D12S391

- Given the close proximity of these loci, should we be concerned when we evaluate the rarity of a match?
- It depends on the question asked…

Syntenic Loci – vWA and D12S391

- For RMP statistics – there is no apparent association between the loci. It is perfectly fine to multiply these probabilities as the loci act independently.
- For very close kinship scenarios (parent-child). The association DOES make a difference.



An evaluation of potential allelic association between the STRs vWA and D12S391: Implications in criminal casework and applications to short pedigrees

Peter Gill [a,b,*], Chris Phillips [c], Catherine McGovern [d], Jo-Anne Bright [d], John Buckleton [d]

[a] Department of Forensic Genetics, Norwegian Institute of Public Health, 0403 Oslo, Norway
[b] Rikshospitalet, University of Oslo, Norway
[c] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Spain
[d] ESR, PB 92021, Auckland, New Zealand

Forensic Science International: Genetics 6 (2012) 477–486

Short communication

Effect of linkage between vWA and D12S391 in kinship analysis☆

Kristen L. O'Connor [a,1], Andreas O. Tillmar [b,*]

[a] U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8314, United States
[b] National Board of Forensic Medicine, Department of Forensic Genetics and Forensic Toxicology, Linköping, Sweden

Forensic Science International: Genetics 6 (2012) 840–844

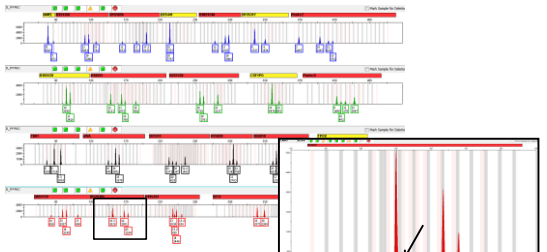
Possible solutions

- (1) Drop one of the loci *(Budowle et al. IJLM (2011))*.
- (2) Treat the two loci as a haplotype.
- (3) Incorporate the recombination frequency into the LR *(Gill et al. FSI:G (2012))*.

One-base Separation

- Noted during the CVP when component D of the NIST SRM 2391c was analyzed.
- D12S3391 – major = 18.3, 22
  minor = 19, 23

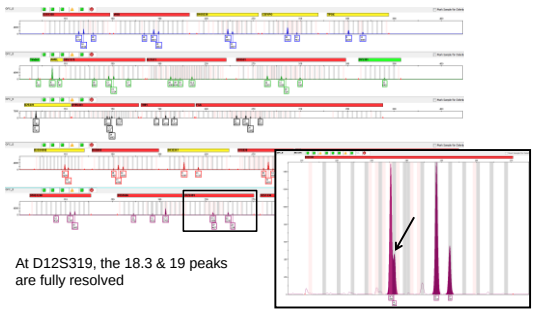
Component D Resolution Issue at 12S391 PowerPlex Fusion



At D12S319, the 18.3 & 19 peaks are not fully resolved so the 19 peak is not called

1 ng DNA, 30 cycles

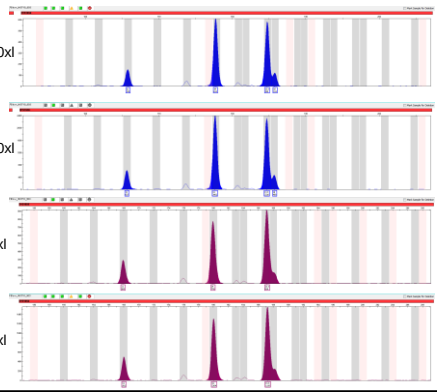
GlobalFiler Component D at D12S391



At D12S319, the 18.3 & 19 peaks are fully resolved

1 ng DNA, 29 cycles

D1S1656



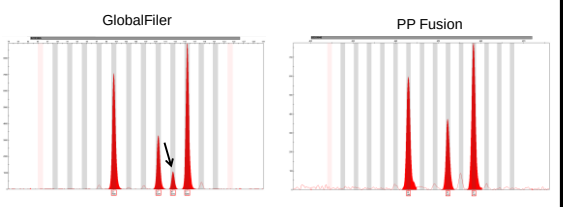
One-base Separation

- This has also been observed in TF kits
- D1S1656 may also show this phenomena (several common microvariants).
- May especially be an issue with mixture samples…

Stutter Issues

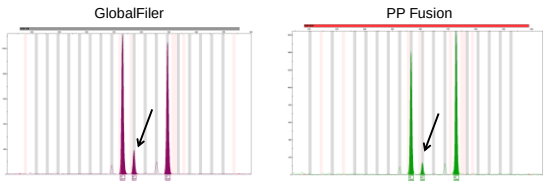
- Some labs are reporting an increase in +4 (forward) stutter.
- Be aware of D22S1045 – trinucleotide repeat.
- Developed as a non-CODIS miniSTR locus.

D22S1045 Issue



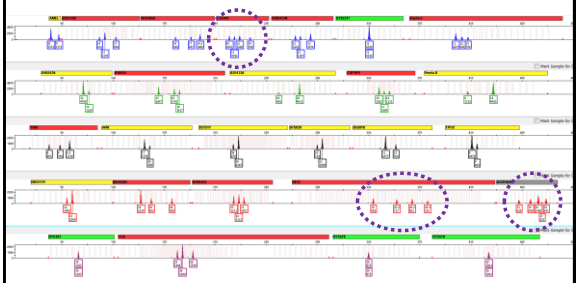
High stutter peak was called with GlobalFiler but not with PP Fusion for Sample 9 (mixt

D2S1338 Issue



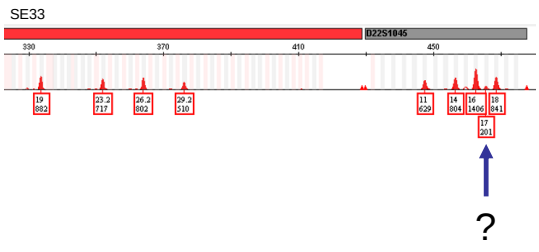
High forward stutter peak was called with both kits for NIST Sample 6

DNA Mixture



9947a+9948 (1:1 ratio)

9947a+9948 (1:1 ratio)

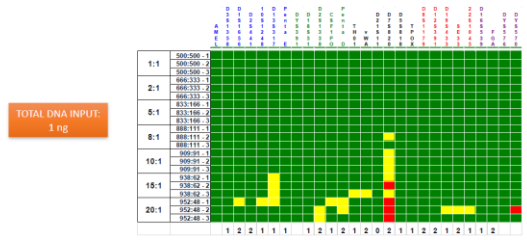


Mixtures

Goals of validating mixtures

- How low can I detect the minor contributor?
- Am I paying attention to major contributor stutter and minor contributor alleles?

DNA mixtures series: Relative ratios of 1:1, 2:1, 5:1, 8:1, 10:1, 15:1, and 20:1 of female / male



Promega Webinars Series: Genetic Identity (05/02/2015) - Validation Studies on DNA Mixtures Using the PowerPlex® Fusion 6C System – Pablo Martin (Spain)

Martin et al. (2014) FSI:G

Goals of validating mixtures

- Higher order mixtures (3+) – you can't test all possible ratios!
- The greater concern with 3+ person mixtures is allele stacking!
- Can you detect all contributors? Does stacking lead to false genotypes?

Forensic Science International: Genetics 19 (2015) 207–211

Short communication

Uncertainty in the number of contributors in the proposed new CODIS set

Michael D. Coble[a,*], Jo-Anne Bright[b,c], John S. Buckleton[a,b], James M. Curran[c]

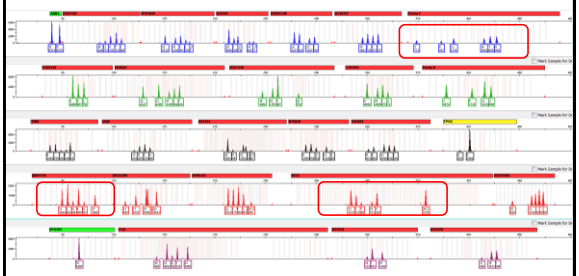
[a] National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[b] ESR, Private Bag 92021, Auckland 1142, New Zealand
[c] University of Auckland Department of Statistics, Private Bag 92019, Auckland 1142, New Zealand

Table 3
Cumulative probability of an N person mixture appearing as a k or fewer person mixture, where $k = 1, \ldots, N-1$ for the Caucasian allele frequencies.

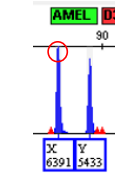
Configuration (# loci)	N contributor mixture	Appearing as k or fewer				
		1	2	3	4	5
Existing CODIS (13)	6	2.13E-40	6.3E-09	0.1612	0.9456	0.9999
GlobalFiler (21)		5.4E-75	9.11E-21	5.31E-05	0.1882	0.8599
Proposed CODIS (20)		1.18E-66	1.66E-16	0.0038	0.6798	0.9969
Fusion (22)		1.51E-75	2.11E-19	0.0009	0.5758	0.9941
Existing CODIS (13)	5	9.66E-33	2.10E-06	0.4141	0.9897	
GlobalFiler (21)		8.21E-61	7.1E-15	0.0048	0.6099	
Proposed CODIS (20)		5.82E-54	9.1E-12	0.0592	0.9228	
Fusion (22)		3.57E-61	7.8E-14	0.0270	0.8885	
Existing CODIS (13)	4	6.82E-25	0.0005	0.7856		
GlobalFiler (21)		4.15E-46	3.5E-09	0.1653		
Proposed CODIS (20)		6.13E-41	3.1E-07	0.4323		
Fusion (22)		2.05E-46	1.8E-08	0.3369		
Existing CODIS (13)	3	8.4E-17	0.0595			
GlobalFiler (21)		5.83E-31	0.0004			
Proposed CODIS (20)		1.66E-27	0.0031			
Fusion (22)		3.52E-31	0.0010			
Existing CODIS (13)	2	1.7E-08				
GlobalFiler (21)		2.1E-15				
Proposed CODIS (20)		1E-13				
Fusion (22)		1.60E-15				

Coble et al. in press

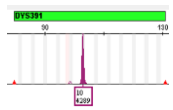
Y-STRs can give guidance



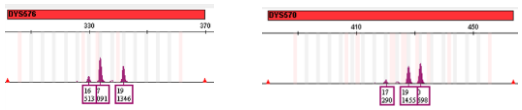
Y-STRs can give guidance



Minor female?



Not Helpful



At least 3 males in the mixture…

Mixtures

- Are you testing the limits of mixture detection?
- If you plan to analyze 4 and 5 person mixtures, VALIDATE this! Don't apply simple rules for a 2-person 8:1 mixture to a 5-person mixture where drop-out is possible.

Goals of Validation

- What are the limitations of the system?
- Where are there failures?
- What can be done to improve or correct?
- What is the actual capacity & capability of the system being validated?

Courtesy of Charlotte Word

From Validation to SOP

Technical SOPs

- Generally this is the easiest part since the procedural steps are often provided by the developer or supplier
  - Often can "cut and paste" into the technical or procedural SOP
  - But must evaluate the validation data to make adjustments
- Did the kits, instruments, etc. perform as expected from developmental validation studies using the technical procedures provided by the developer or supplier?

Charlotte J Word - 2015

Courtesy of Charlotte Word

Technical SOPs

- Critical to evaluate all parameters tested during the validation
  - What are the ranges of the assay where "good" data are generated?
  - What are the "edges" of the system?
  - Is more testing needed to define "the edges?"
  - What are the limitations of the assay?
- May have to modify the provided procedures
- Important to define the testing parameters ranges that MUST be followed

Charlotte J Word - 2015

Courtesy of Charlotte Word

Interpretation SOPs

- This is an area needing MUCH attention at this time
- Important to recognize that interpretation guidelines and SOPs MUST come from the data generated in the validation studies
  - Some guidance from published data and recommendations (e.g., ISFG, SWGDAM)

Charlotte J Word - 2015

Courtesy of Charlotte Word

Interpretation SOPs

- Do the validation study samples tested reflect all of the types of samples accepted and tested in the laboratory?
- Do the interpretation SOPs cover all types of samples accepted and tested in the laboratory?
- Can't make interpretation procedures regarding low template samples or complex mixtures without having validation data to evaluate
  - and needed for training!

Charlotte J Word - 2015

Courtesy of Charlotte Word

Interpretation SOPs

- Important to determine:
  - What are the limitations of the assay?
    - Under what conditions…
      - Are correct full results obtained?
      - Are the data reproducible in your laboratory and by another lab?
      - Can the results lead to possible misinterpretations due to partial profiles or additional data?
  - Where and what cautions are needed
  - When do the results become uninterpretable or inconclusive?

Charlotte J Word - 2015

Courtesy of Charlotte Word

Validation of the Interpretation SOPs

It is important to demonstrate that the interpretation SOPs:

- Generate correct and accurate conclusions
- Are detailed enough to provide consistency within the laboratory
  - Do all analysts report the same alleles and genotypes from the same data?
    - How are decisions made regarding artifacts? Thresholds? What flexibility exists?
  - Do all analysts get the same conclusions from the same data?

Charlotte J Word - 2015

Courtesy of Charlotte Word

Validation of Interpretation SOPs

- If inconsistency within the laboratory:
  - SOPs not detailed or clear enough
  - SOPs do not cover all needed parameters or scenarios adequately
  - Have possibly identified types of samples that should not be interpreted in your laboratory

Charlotte J Word - 2015

Courtesy of Charlotte Word

Implementation of New SOPs

- New QA/QC measures
  - Incorporate into proficiency test cycle
  - Critical reagents or instruments
  - Periodic monitoring/assessment
- Training of Analysts
  - May be short or long depending on what is being introduced
  - Competency test
- Assessment for QAS and accreditation at next external audit/inspection

Charlotte J Word - 2015

Courtesy of Charlotte Word

Implementation of New SOPs

- Additional training/notification of availability
  - Law enforcement
  - Attorneys – prosecution and defense
  - Judges
- Discovery
  - New SOPs (some laboratories post on internet)
  - Validation studies supporting the new SOPs
    - Summaries, tables/graphs of results, actual data
  - List of publications supporting the new assay
  - Training files

Charlotte J Word - 2015

Courtesy of Charlotte Word

Implementation of New SOPs

- Consider if changes are needed to case acceptance policies
  - Are these new procedures suited to all types of samples currently accepted?
  - Need to decrease acceptance of any types?
  - Can you expand the types of samples accepted?
- Consider impact on previous cases
  - Can additional studies now be done?
  - Do the new studies invalidate anything done previously?

Charlotte J Word - 2015

Courtesy of Charlotte Word

Implementation of New System

- Plan and prepare for possible admissibility hearing
  - Publications
  - SOPs
  - Validation studies
- May need guidance from experienced attorneys and scientists
  - Very different from routine trial testimony

Charlotte J Word - 2015

Courtesy of Charlotte Word

Planning for the next validation

- Use the same samples you have tested for your old STR multiplex for the next STR multiplex.
- Positive controls, current employees, and NIST SRM components are popular examples, but have limitations.

Using Analysts as Samples…

Genetic Information Nondiscrimination Act of 2008

Long title	An act to prohibit discrimination on the basis of genetic information with respect to health insurance and employment.
Acronyms (colloquial)	GINA
Enacted by	the 110th United States Congress
Effective	May 21, 2008

Planning for the next validation

- Consider – buying a large supply of blood (anonymously collected) from a blood bank.
- Characterize as much as you can (nice summer intern project).
- Why?

FSI-Genetics 5(5):376-80

Analysis and interpretation of mixed profiles generated by 34 cycle SGM Plus® amplification

Jon H. Wetton *, John Lee-Edghill, Emily Archer, Valerie C. Tucker, Andrew J. Hopwood, Jonathan Whitaker, Gillian Tully

Forensic Science Service, 2960 Trident Court, Birmingham Business Park, Solihull B37 7YN, UK

Table 2
The final distribution of major and consensus profiles across the template input and ratio range after completion of the RO requested rework.

Ratio	Total input	Major/minor	Major/ consensus	Consensus only	Unduplicated
5:1	1 ng	12		3	1
	500 pg	10[a]		6	
	250 pg	11[a]	1	4	
	100 pg	3[a]	4	9	
	50 pg	2		14	
	Total	38	5	36	1
2:1	1 ng	1		14	1
	500 pg	1[a]		14	1
	250 pg	1[a]		15	
	100 pg		1	14	1
	50 pg		1[a]	15	
	Total	3	2	72	3
1:1	1 ng			8	
	500 pg			8	
	250 pg			8	
	100 pg			8	
	50 pg		1[a]	7	
	Total	0	1	39	0

[a] One observation in each of these major profile categories was due to a single individual with a possible mutation affecting the accuracy of the Quantifiler® concentration estimate which may have caused the input of this individual to be underestimated by half. This would tend to double its relative representation at each mixture ratio as well as the true amount of template available in the PCR. In the other rows the affected mixture was scored as a consensus.

One individual had a possible mutation in the Quantifiler primer binding region that underestimated the relative ratios in the study…

Planning for the next validation

- Don't be an island – consider joining with other labs to plan validation studies. Other labs may have ideas that can streamline your validation.
- Share resources, cost, time. Perhaps you may find a "Steven Myers" type who saves your lab hours with a neat Excel spreadsheet.
- Inter-laboratory challenges to monitor progress.

No matter what you are validating now…

- This is NOT your last validation!
- Should you be an early adopter?
- Should you go kicking and screaming?

Here now…

Coming very soon…

NetBio ANDE

IntegenX RapidHIT 200

Coming later…

Ion Torrent

MiSeq

Final thoughts

- Proper validation is hard work. It takes a great deal of time and effort to conduct the studies, write the new SOPs, train everyone, and monitor the progress.
- A good plan up-front will go a long way to make this process as easy as possible.

Acknowledgments

Dr. Charlotte Word
Dr. Robin Cotton
Dr. Jo Bright

Contact info:
mcoble@nist.gov
+1-301-975-4330