Consortium Validation Project Data Assessment to Finalize the Expanded CODIS Core Loci in the U.S.

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

Carolyn R. Steffen, Margaret C. Kline, David L. Duewer, Peter M. Vallone, and Douglas R. Hares

U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA

Email: Becky.Steffen@nist.gov

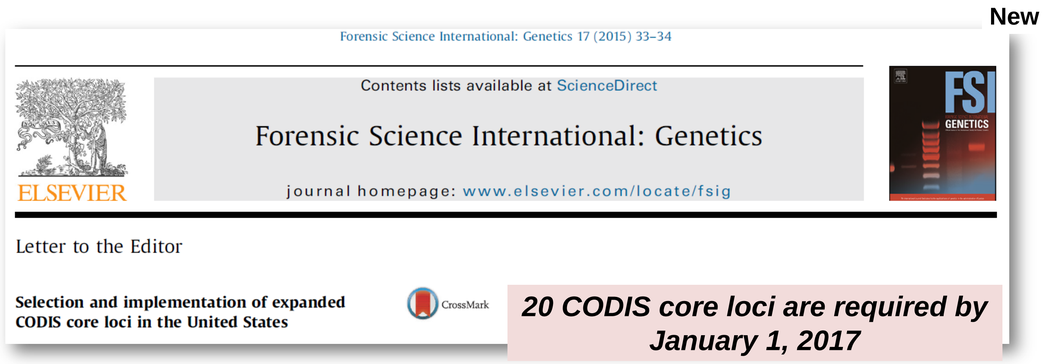
ISFG

P-241

In 2012 the FBI proposed to expand the number of CODIS core loci in the United States [1,2]. With the development of commercial STR multiplex kits containing the expanded loci, a consortium of 11 CODIS laboratories were selected to evaluate the performance of these kits in a collaborative validation study [3]. The study examined data from a variety of sample types including single source known, non-probative, challenge, and mixed samples. The study assessed concordance, precision, reproducibility, sensitivity, and mixture detection. The results were compiled and summarized using software developed at NIST to assist in the selection and adoption of the expanded CODIS core loci in the U.S. This presentation explores a number of aspects of the data analysis including stutter and heterozygote balance results within the kits and lessons learned regarding the markers that led to final core loci. This presentation documents the data driven decisions made to expand the CODIS core Loci in the U.S.

References: 1. Hares D: Expanding the CODIS core loci in the United States. FSI Genet. 2012; 6: e52-e54 2. Hares D: Addendum to expanding the CODIS Core Loci in the United States, FSI Genet. 2012; 6: e135 3. Hares D: Selection and Implementation of Expanded CODIS Core Loci in the United States. FSI Genet. 2015; 17: 33-34

Selection of the 20 expanded CODIS core loci in the United States was announced in November 2014 and implementation of these loci is required by January 1, 2017


Forensic Science International: Genetics 17 (2015) 33–34

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

ELSEVIER

Letter to the Editor

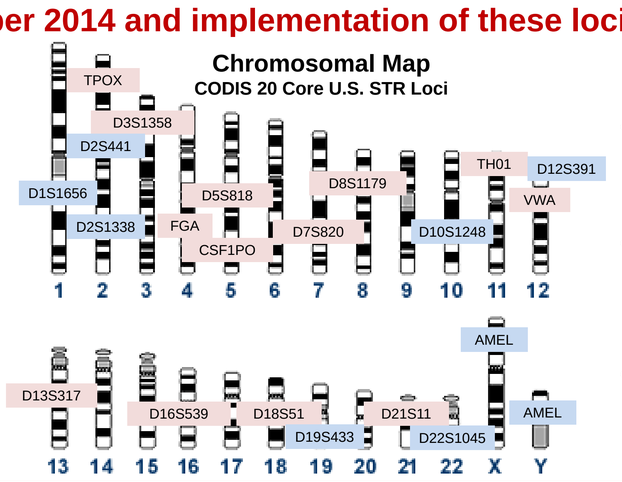
Selection and implementation of expanded CODIS core loci in the United States

CrossMark

20 CODIS core loci are required by January 1, 2017

New 20 CODIS Core Loci

Locus	
CSF1PO	Original CODIS Core 13 Loci (red)
D3S1358	
D5S818	
D7S820	
D8S1179	
D13S317	
D16S539	
D18S51	
D21S11	
FGA	
TH01	
TPOX	
vWA	
D1S1656	New additional CODIS Core Loci (blue)
D2S441	
D2S1338	
D10S1248	
D12S391	
D19S433	
D22S1045	



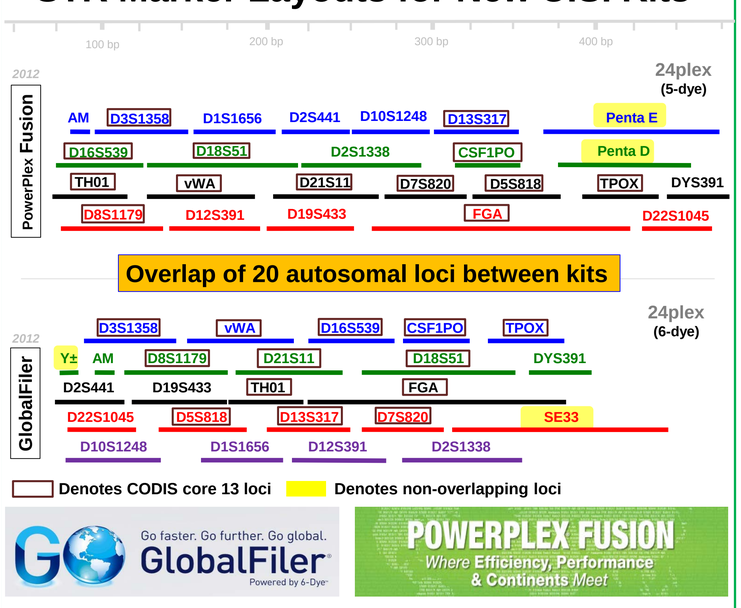
FBI Consortium Validation Project (CVP)
October 2012 – July 2014

- 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

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

STR Marker Layouts for New U.S. Kits



Purpose of the FBI CVP

- To have several crime labs across the U.S. assess two new multiplex kits that include the new loci
- To determine if these new loci meet the criteria to become part of the CODIS core set
- To facilitate implementation of the new loci into the U.S. National DNA Database

Implementation Plan

- Plan Published in February 2012

http://www.fbi.gov/about-us/lab/codis/planned-process-and-timeline-for-implementation-of-additional-codis-core-loci

- Selection of laboratories to participate in validation studies
- Validation of proposed new CODIS core loci
- Selection of CODIS core loci
- Implementation of new CODIS core loci into NDIS operations

Validation Project Participants

Laboratory	
Armed Forces DNA Identification Laboratory (AFDIL)	Missing Person Casework
California Department of Justice (CalDOJ)	Database and Casework (including missing persons)
Federal Bureau of Investigation (FBI)	Database and Casework (including missing persons)
Indiana State Police (INSP)	Database and Casework
Illinois State Police (ISP)	Database and Casework
Louisiana State Police (LSP)	Database
Missouri State Highway Patrol (MSHP)	Database and Casework
National Institute of Standards and Technology (NIST)	Data evaluation
Ohio Bureau of Criminal investigation (OBCI)	Database and Casework
Palm Beach County Sherriff's Office (PBSO)	Casework
Santa Clara County DA (SCC)	Casework
Virginia Department of Forensic Sciences (VADFS)	Database and Casework

Validation Plan

- Some key areas of interest (performance of kits):
  - Genotype concordance
  - Sensitivity
  - Precision
  - Non-Allelic PCR products
  - Heterozygous peak signal balance
  - Stutter
  - Stochastic PCR thresholds
  - Mixtures
  - Casework challenge samples

Consortium Validation Project (CVP) Data Analysis

CVP 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 sets)	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		

Component 1 of the CVP

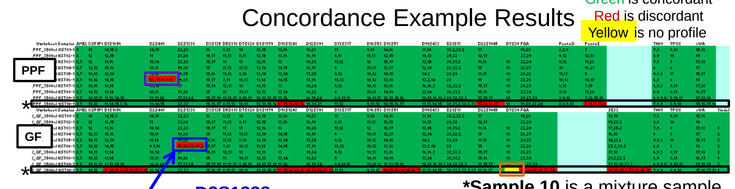
NIST Known Sample Study

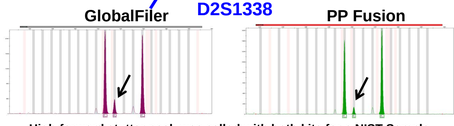
- NIST provided 10 samples with DNA extract on FTA and 903 paper
- There were 8 single source samples and 2 mixture samples
- Concordance, peak height ratios, and stutter percentages were all evaluated from this data
- NIST samples were some of the first samples run with the new kits
- Mixture samples highlighted areas to explore further in the validation effort
- Laboratories were able to reevaluate any highlighted data, concordance was achieved

NIST Known Sample Study Data

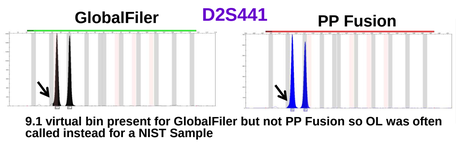
Green is concordant
Red is discordant
Yellow is no profile

Concordance Example Results





High forward stutter peak was called with both kits for a NIST Sample This is actually a sample artifact



9.1 virtual bin present for GlobalFiler but not PP Fusion so OL was often called instead for a NIST Sample

***Sample 10** is a mixture sample
DYS391 is full locus dropout

Peak Height Ratios (Median)		
	GlobalFiler	PP Fusion
3130xl	90.4%	91.3%
3500xl	90.3%	89.2%

Stutter Percentage (Median)		
	GlobalFiler	PP Fusion
3130xl	5.93%	5.80%
3500xl	5.94%	5.76%

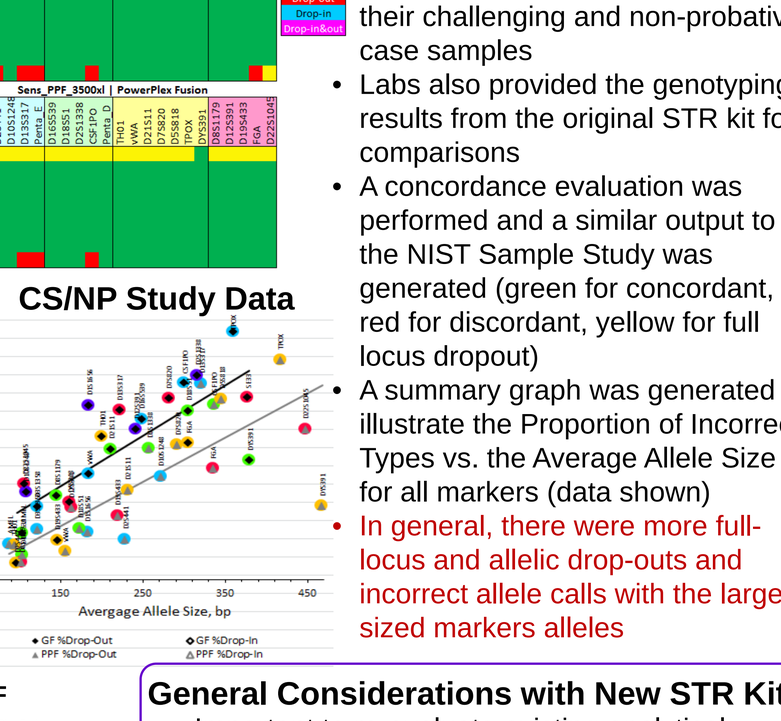
Known/Non-Probative Study

- Labs provided data genotyped with the two kits from their known and non-probative samples
- Labs also provided the genotyping results from the original STR kit for comparisons
- A concordance evaluation was performed and a similar output to the NIST Sample Study was generated (green for concordant, red for discordant, yellow for full locus dropout)
- In general, concordance was high for both kits (data not shown)
- Laboratories were able to reevaluate any highlighted data, concordance was achieved

Precision Study

- The participating labs used the instrumentation present in their laboratories
- Many instruments were used including 3130xl and 3500xl
- Labs were instructed to run a full injection of allelic ladders to evaluate precision using both kits
- In general, precision was consistent for both kits
- Allele Size Standard Deviation ranges from 0.036 – 0.076 basepairs (bp)
- Some labs used different data smoothing (i.e. Local Southern vs 3rd Order) – all data was evaluated, only some example data is shown

Precision Study Data

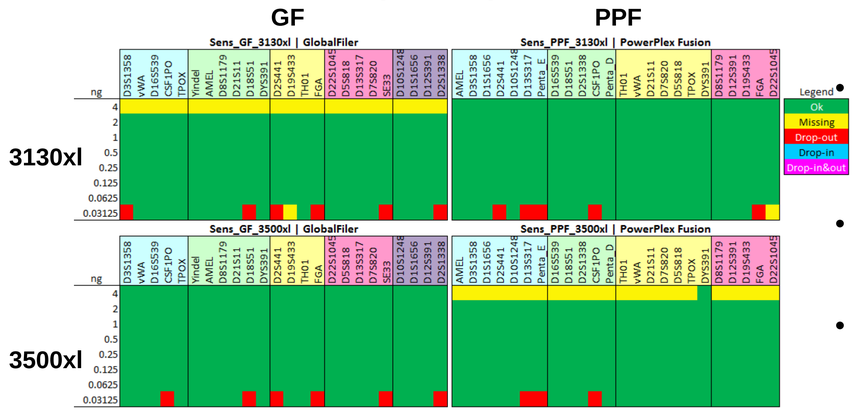


Component 2 of the CVP

Sensitivity Study

- Labs provided data run with two replicates of each dilution series (4, 2, 1, 0.5, 0.25, 0.125, 0.06, and 0.03 ng of total DNA in amp)
- Consensus profiles were used to determine the correct genotypes for the two new kits
- Heat maps summarize the results of the data evaluation
- In general, 4 ng was too much DNA and often resulted in blown out signal and drop-out was observed starting at 0.5 ng (data not shown) down to 0.03 ng (data shown)

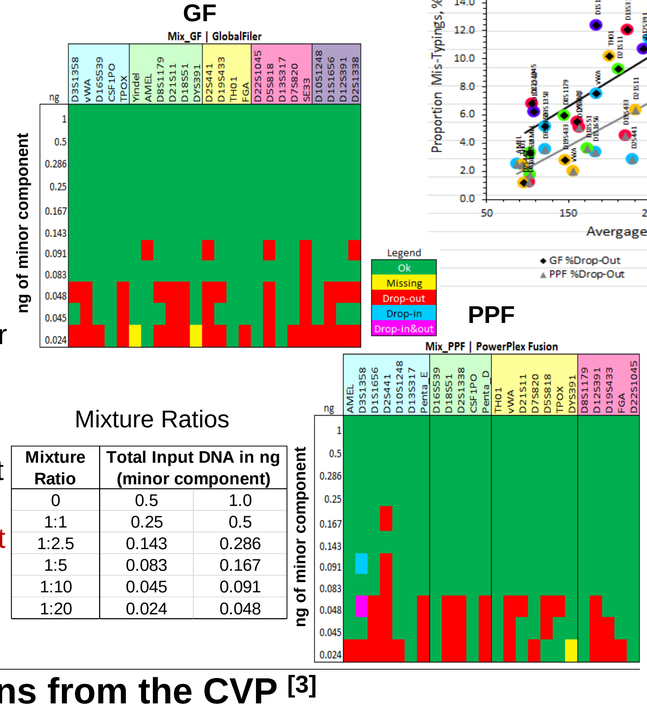
Sensitivity Study Data



Mixture Study

- Labs provided data run with 2 person (example data shown) and 3 person mixtures to include 2 replicates with 1 ng total DNA in amp with the following ratios: 1:10, 1:5, 1:2.5, 1:1, 2.5:1, 5:1, 10:1, 0
- Consensus profiles were used to determine the correct genotypes for the two new kits
- Heat maps summarize the results of the data evaluation (ratios were converted to total minor component DNA (in ng) present in the mixture)
- In general, as the minor component concentration decreased, more drop-out was observed

Mixture Study Data



Mixture Ratios

Mixture Ratio	Total Input DNA in ng (minor component)	
0	0.5	1.0
1:1	0.25	0.5
1:2.5	0.143	0.286
1:5	0.083	0.167
1:10	0.045	0.091
1:20	0.024	0.048

CS/NP Study Data

Challenging Samples/ Non-Probative Case (CS/NP) Material Study

- Labs provided genotypes generated with the two kits from their challenging and non-probative case samples
- Labs also provided the genotyping results from the original STR kit for comparisons
- A concordance evaluation was performed and a similar output to the NIST Sample Study was generated (green for concordant, red for discordant, yellow for full locus dropout)
- A summary graph was generated to illustrate the Proportion of Incorrect Types vs. the Average Allele Size for all markers (data shown)
- In general, there were more full-locus and allelic drop-outs and incorrect allele calls with the larger-sized markers alleles

Conclusions from the CVP [3]

- 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, and SE33 (also present in commercial STR kits) are accepted loci but are ***NOT*** required CODIS Core Loci

General Considerations with New STR Kits

- Important to re-evaluate existing analytical thresholds, stochastic thresholds and interpretation guidelines during validation and implementation of these new multiplex kits.
- New kits are very sensitive
  - May see increase in mixtures
  - May see increase in contamination
- Manufacturers recommendations for cycle number is higher
  - 30 cycles for Fusion
  - 29 cycles for Globalfiler
  - Consider incorporating a cycle number study into your validation – may be able to decrease a cycle or two and still achieve desired sensitivity

Poster available for download from STRBase
http://www.cstl.nist.gov/biotech/strbase/pub_pres/SteffenISFG2015CVPposter.pdf