NIST Experience with FSS-i^3 Software

Becky Hill and Amy Decker
Margaret C. Kline and John M. Butler

National Institute of Standards and Technology

Promega Summer Technology Tour – June 2006

Disclaimers

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

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

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

Our publications and presentations are made available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

Packaging for FSS *i-Cubed* Software

Currently very limited documentation is provided with the software (some PowerPoint files are on CD-ROM)

Cost to the End User

- **Software**
  - single copy, single computer **$20,000**
  - If 5 copies purchased, then $50,000
- **Maintenance agreement**
  - **$4,000** per year
  - Software upgrades are included
- **Training**
  - **$2,000** if at Promega (plus your travel expenses)
  - $12,000 for up to 5 people if performed in your lab
- Requires GeneMapper ID or GeneScan/Genotyper software to already be in place in your lab

Minimum starting cost of $26,000

NIST Experience with Software Purchase

- Attempted to purchase directly from FSS
  - No quote provided by Chris Macguire despite multiple attempts and email agreements to do so (Dec 2004, Jan, Feb, May, June 2005)
- Quote for software from Promega on Oct 18, 2005
  - Told that we had to purchase $4,000 maintenance agreement along with at least $2,000 training (plus travel expense to Madison, WI)
  - NIST contract officer signed off Dec 19, 2005
- Promega installed software January 3, 2006
- Becky Hill went to Madison, WI Jan 9-13, 2006 for first training class held at Promega

Promega Training Overview

- Day 1: Introduction to the software, batching of data (macro), creating RAW files from GenoTyper and GeneMapper ID, input templates
- Day 2: Scientific settings, ladder templates, FSS-i^3 Rule Sets
- Day 3: Settings folder, output templates, i-ntegrity Module
- Day 4: Mixture Interpretation Theory of i-STReam Module
- Day 5: Review of software features

Agreements Coming with Software

- There are two documents imbedded in the software installation that must be reviewed and accepted prior to loading FSS-i^3 software.
- Hard copies were not provided and had to be obtained later from Promega.

Overview of Software Components

i-STRess quickly and accurately calls your allele types and objectively assesses the quality of your data. This allows analysts to reduce the time spent manually reviewing data and focus on "problem" samples.

i-STReam module evaluates two-person DNA mixtures and produces a best-fit major profile. This aids the reporting analyst in mixture deconvolution and unbiased interpretation.

i-ntegrity checks for potential sample-to-sample contamination within a batch by comparing all alleles called in a sample to the alleles in every other sample in the batch.

http://www.promega.com/applications/profile.asp?appname=Genetic+Identity&sku=FSSI3&spl=off

Features of FSS-i^3

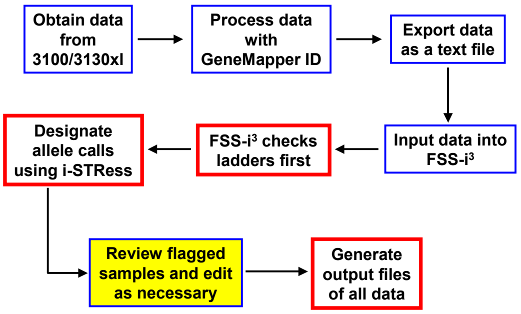
listed on Promega website

http://www.promega.com/applications/profile.asp?appname=Genetic+Identity&sku=FSSI3&spl=off

- **Integrates with existing commercial software**
- **Works with most autosomal STR multiplex kits**
- **User-customizable input and output files**
- **User-configurable settings for optimized data analysis**
- **Two-person DNA mixture deconvolution**
- **Contamination Check**

Introduction to FSS-i^3 Software

FSS-i^3 Flow Chart



Obtain data from 3100/3130xl

Process data with GeneMapper ID

Export data as a text file

Input data into FSS-i^3

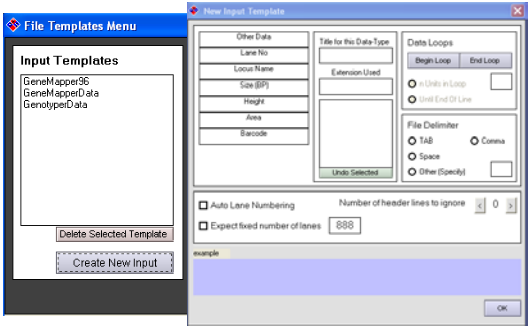
FSS-i^3 checks ladders first

Designate allele calls using i-STRess

Review flagged samples and edit as necessary

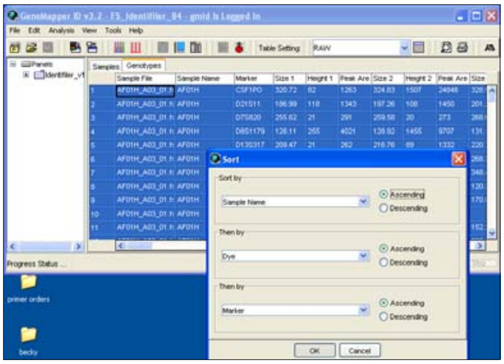
Generate output files of all data

FSS-i^3 Input Template



Controls what and how information will be brought into the FSS-i^3 program

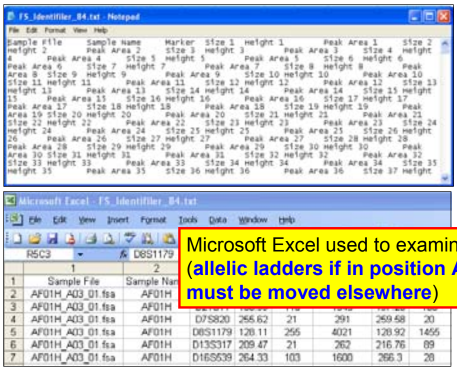
GeneMapper ID Raw Data Sorted



	Sample File	Sample Name	Marker	Size 1	Height 1	Peak Are	Size 2	Height 2	Peak Are	Size
1	AF01H_A03_01.f:	AF01H	CSF1PO	320.72	82	1263	324.83	1507	24848	328
2	AF01H_A03_01.f:	AF01H	D21S11	186.99	118	1343	187.26	108	1450	201
3	AF01H_A03_01.f:	AF01H	D7S820	255.62	21	291	259.58	20	273	268
4	AF01H_A03_01.f:	AF01H	D8S1179	128.11	255	4021	128.92	1455	9707	131
5	AF01H_A03_01.f:	AF01H	D13S317	209.47	21	262	216.76	89	1332	220
6	AF01H_A03_01.f:	AF01H								268
7	AF01H_A03_01.f:	AF01H								346
8	AF01H_A03_01.f:	AF01H								120
9	AF01H_A03_01.f:	AF01H								170
10	AF01H_A03_01.f:	AF01H								
11	AF01H_A03_01.f:	AF01H								152

Sort by: Sample Name ◉ Ascending ○ Descending
Then by: Dye ◉ Ascending ○ Descending
Then by: Marker ◉ Ascending ○ Descending

Text File Saved from GeneMapper ID "Raw Data"

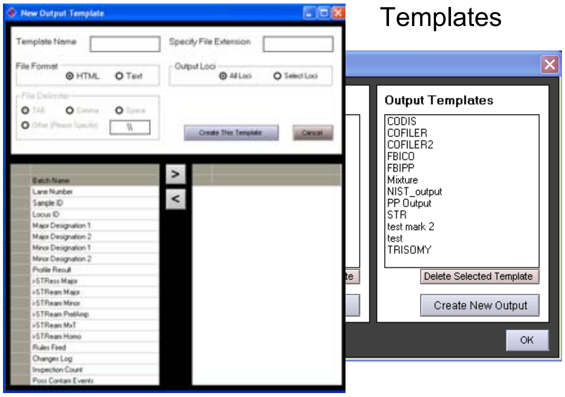


```
Sample File     Sample Name     Marker  Size 1  Height 1        Peak Area 1     Size 2
Height 2        Peak Area 2     Size 3  Height 3        Peak Area 3     Size 4  Height
4       Peak Area 4     Size 5  Height 5        Peak Area 5     Size 6  Height 6
Peak Area 6     Size 7  Height 7        Peak Area 7     Size 8  Height 8        Peak
Area 8  Size 9  Height 9        Peak Area 9     Size 10 Height 10       Peak Area 10
Size 11 Height 11       Peak Area 11    Size 12 Height 12       Peak Area 12    Size 13
Height 13       Peak Area 13    Size 14 Height 14       Peak Area 14    Size 15 Height
15      Peak Area 15    Size 16 Height 16       Peak Area 16    Size 17 Height 17
Peak Area 17    Size 18 Height 18       Peak Area 18    Size 19 Height 19       Peak
Area 19 Size 20 Height 20       Peak Area 20    Size 21 Height 21       Peak Area 21
Size 22 Height 22       Peak Area 22    Size 23 Height 23       Peak Area 23    Size 24
Height 24       Peak Area 24    Size 25 Height 25       Peak Area 25    Size 26 Height
26      Peak Area 26    Size 27 Height 27       Peak Area 27    Size 28 Height 28
Peak Area 28    Size 29 Height 29       Peak Area 29    Size 30 Height 30       Peak
Area 30 Size 31 Height 31       Peak Area 31    Size 32 Height 32       Peak Area 32
Size 33 Height 33       Peak Area 33    Size 34 Height 34       Peak Area 34    Size 35
Height 35       Peak Area 35    Size 36 Height 36       Peak Area 36    Size 37 Height
```

	1	2						
1	Sample File	Sample Nan						
2	AF01H_A03_01.fsa	AF01H						
3	AF01H_A03_01.fsa	AF01H						
4	AF01H_A03_01.fsa	AF01H	D7S820	255.62	21	291	259.58	20
5	AF01H_A03_01.fsa	AF01H	D8S1179	128.11	255	4021	128.92	1455
6	AF01H_A03_01.fsa	AF01H	D13S317	209.47	21	262	216.76	89
7	AF01H_A03_01.fsa	AF01H	D16S539	264.33	103	1600	266.3	28

Microsoft Excel used to examine data (**allelic ladders if in position A01 must be moved elsewhere**)

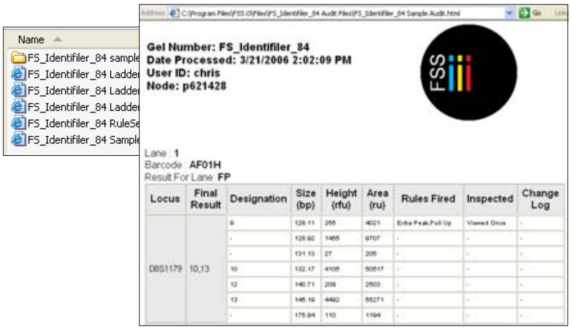
FSS-i^3 Output Templates



File Format: ◉ HTML ○ Text
Output Loci: ◉ All Loci ○ Select Loci

Output Files Created by FSS-i^3 i-STRess

C:\Program Files\FSS i3\Files\FS_Identifiler_84 Audit Files



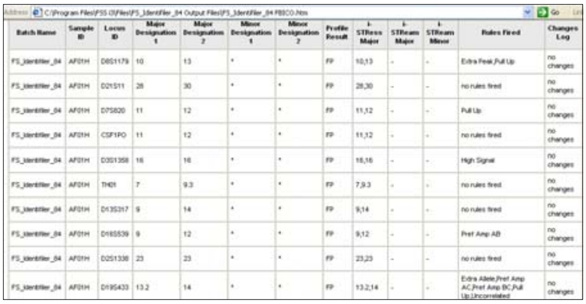
Gel Number: FS_Identifiler_84
Date Processed: 3/21/2006 2:02:09 PM
User ID: chris
Node: p621428

Lane : **1**
Barcode : **AF01H**
Result For Lane: **FP**

Locus	Final Result	Designation	Size (bp)	Height (rfu)	Area (ru)	Rules Fired	Inspected	Change Log
D8S1179	10,13	8	128.11	255	4021	Extra Peak,Pull Up	Viewed Once	-
		-	128.92	1455	9707	-	-	-
		-	131.13	27	205	-	-	-
		10	132.17	4106	60517	-	-	-
		12	140.71	209	2503	-	-	-
		13	146.19	4492	55271	-	-	-
		-	175.94	110	1194	-	-	-

Output Files Created by FSS-i^3 i-STRess

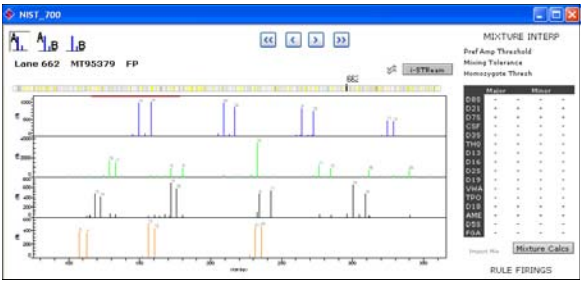
C:\Program Files\FSS i3\Files\FS_Identifiler_84 Output Files



Batch Name	Sample ID	Locus ID	Major Designation 1	Major Designation 2	Minor Designation 1	Minor Designation 2	Profile Result	i-STRess Major	i-STReam Major	i-STReam Minor	Rules Fired	Changes Log
FS_Identifiler_84	AF01H	D8S1179	10	13	*	*	FP	10,13	-	-	Extra Peak,Pull Up	no changes
FS_Identifiler_84	AF01H	D21S11	28	30	*	*	FP	28,30	-	-	no rules fired	no changes
FS_Identifiler_84	AF01H	D7S820	11	12	*	*	FP	11,12	-	-	Pull Up	no changes
FS_Identifiler_84	AF01H	CSF1PO	11	12	*	*	FP	11,12	-	-	no rules fired	no changes
FS_Identifiler_84	AF01H	D3S1358	16	16	*	*	FP	16,16	-	-	High Signal	no changes
FS_Identifiler_84	AF01H	TH01	7	9.3	*	*	FP	7,9.3	-	-	no rules fired	no changes
FS_Identifiler_84	AF01H	D13S317	9	14	*	*	FP	9,14	-	-	no rules fired	no changes
FS_Identifiler_84	AF01H	D16S539	9	12	*	*	FP	9,12	-	-	Pref Amp AB	no changes
FS_Identifiler_84	AF01H	D2S1338	23	23	*	*	FP	23,23	-	-	no rules fired	no changes
FS_Identifiler_84	AF01H	D19S433	13.2	14	*	*	FP	13.2,14	-	-	Extra Allele,Pref Amp AC,Pref Amp BC,Pull Up,Uncorrelated	no changes

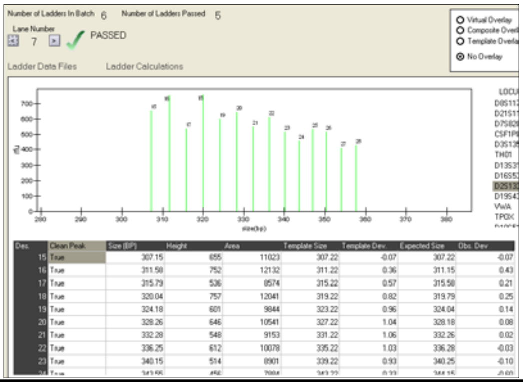
Files are created in html but can easily be imported into Excel for sorting and review

Images from FSS-i^3



Samples are displayed as "spikograms" rather than peaks using the peak size and height information gathered from GeneMapper ID or Genotyper "raw data"

Identifiler D2S1338 Allelic Ladder



Des.	Clean Peak	Size (BP)	Height	Area	Template Size	Template Dev.	Expected Size	Obs. Dev
15	True	307.15	655	11023	307.22	-0.07	307.22	-0.07
16	True	311.58	752	12132	311.22	0.36	311.15	0.43
17	True	315.79	536	8574	315.22	0.57	315.58	0.21
18	True	320.04	757	12041	319.22	0.82	319.79	0.25
19	True	324.18	601	9844	323.22	0.96	324.04	0.14
20	True	328.26	646	10541	327.22	1.04	328.18	0.08
21	True	332.28	548	9153	331.22	1.06	332.26	0.02
22	True	336.25	612	10078	335.22	1.03	336.28	-0.03
23	True	340.15	514	8901	339.22	0.93	340.25	-0.10

Work Performed at NIST

Allele Concordance Studies at NIST

- Manual calls
  - with GeneScan/Genotyper v3.7
  - with GeneMapper ID v3.2
- Automated calls with GM/FSS-i^3
- Comparison of output with Excel spreadsheets written by Dave Duewer (NIST)

NIST Data Sets Available for Examination

"Database" Samples

- Identifiler (ABI 3100 – data collection 1.0.1, POP-6)
  - **700 NIST U.S. population samples** (JFS 2003,48:908-911)
  - 375 cell-line samples from collaborator
  - 500 father-son samples from paternity testing lab
- Identifiler (ABI 3130xl – data collection 3.0, POP-7)
  - 375 cell-line samples from collaborator
- Profiler Plus (ABI 3100 and ABI 3130xl)
  - 95 father-son samples from paternity testing lab
- PowerPlex 16 (ABI 3100 – data collection 1.0.1, POP-6)
  - 318 aged blood stains from collaborator showing degraded profiles
- ProfilerPlus/COfiler and Profiler (ABI 310, POP-4)
  - >3,500 samples from 1998-2001 AFDIL QC sample checks

~5,500 samples processed with commonly used STR kits and instruments

Single Source Samples Examined with i-STRess

FSS i-STRess

- Ran **262 Identifiler** samples (father-son samples) with GM/FSS-i^3 and compared to GeneScan/ Genotyper and GeneMapper ID results
- Ran **864 Identifiler** samples (700 reported NIST U.S. population samples) with GM/FSS-i^3 and compared to GeneScan/Genotyper results (see http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/JFS2003IDresults.xls)

Over 1,200 unique samples will be examined eventually

Issues with Review of Previous Data

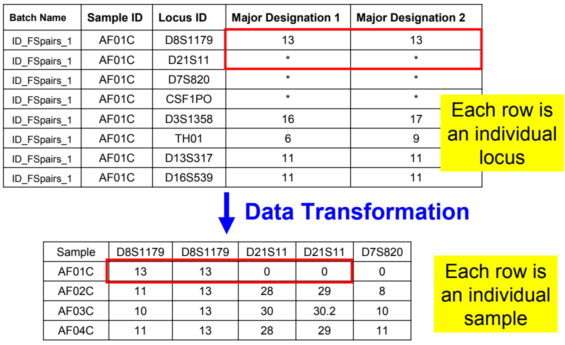
- Need a rapid way to compare allele calls for concordance purposes
  - Allele calls from Genotyper and GeneMapper are in different format from FSS-i^3 output
- **Potential of finding mistakes in original allele calls that you thought were without error**
  - Do you have a protocol for fixing “mistakes”?
  - Error rate in double manual data review is not zero!

Data Comparison Between Methods

- Dave Duewer (NIST Analytical Chemistry Division) has written several computer programs to convert and compare FSS-i^3 data that utilize Excel macros
  - DNA_FSSi3_Convert.xls (converts data format)
  - STR_MatchSamples.xls (compares samples)
- These programs will be made available to the community after additional testing and refinement
  - **http://www.cstl.nist.gov/biotech/strbase/software.htm**

DNA_FSSi3_Convert.xls

First five columns in FSS-i^3 output are converted to be like Genotyper allele designation table



Batch Name	Sample ID	Locus ID	Major Designation 1	Major Designation 2
ID_FSpairs_1	AF01C	D8S1179	13	13
ID_FSpairs_1	AF01C	D21S11	*	*
ID_FSpairs_1	AF01C	D7S820	*	*
ID_FSpairs_1	AF01C	CSF1PO	*	*
ID_FSpairs_1	AF01C	D3S1358	16	17
ID_FSpairs_1	AF01C	TH01	6	9
ID_FSpairs_1	AF01C	D13S317	11	11
ID_FSpairs_1	AF01C	D16S539	11	11

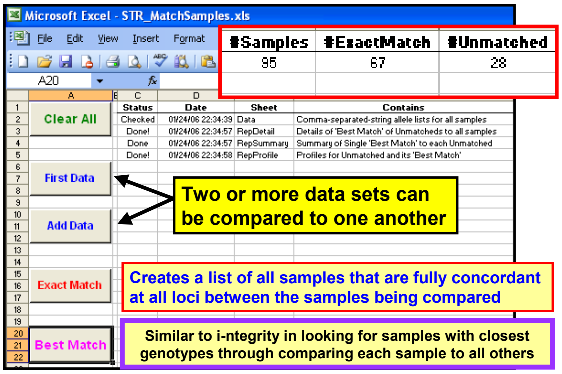
Each row is an individual locus

Data Transformation

Sample	D8S1179	D8S1179	D21S11	D21S11	D7S820
AF01C	13	13	0	0	0
AF02C	11	13	28	29	8
AF03C	10	13	30	30.2	10
AF04C	11	13	28	29	11

Each row is an individual sample

STR_MatchSamples.xls



#Samples	#ExactMatch	#Unmatched
95	67	28

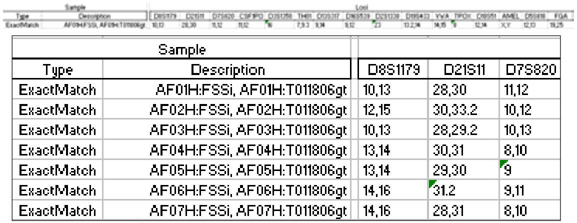
Status	Date	Sheet	Contains
Checked	01/24/06 22:34:39	Data	Comma-separated-string allele lists for all samples
Done!	01/24/06 22:34:57	RepDetail	Details of 'Best Match' of Unmatcheds to all samples
Done	01/24/06 22:34:57	RepSummary	Summary of Single 'Best Match' to each Unmatched
Done!	01/24/06 22:34:58	RepProfile	Profiles for Unmatched and its 'Best Match'

Two or more data sets can be compared to one another

Creates a list of all samples that are fully concordant at all loci between the samples being compared

Similar to i-ntegrity in looking for samples with closest genotypes through comparing each sample to all others

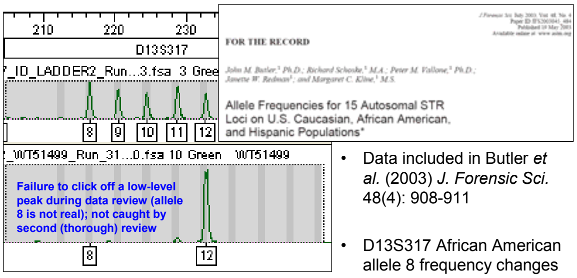
Exact Matches (Full Concordance) Observed with STR_MatchSamples.xls Program



Sample				
Type	Description	D8S1179	D21S11	D7S820
ExactMatch	AF01H:FSSi, AF01H:T011806gt	10,13	28,30	11,12
ExactMatch	AF02H:FSSi, AF02H:T011806gt	12,15	30,33.2	10,12
ExactMatch	AF03H:FSSi, AF03H:T011806gt	10,13	28,29.2	10,13
ExactMatch	AF04H:FSSi, AF04H:T011806gt	13,14	30,31	8,10
ExactMatch	AF05H:FSSi, AF05H:T011806gt	13,14	29,30	9
ExactMatch	AF06H:FSSi, AF06H:T011806gt	14,16	31.2	9,11
ExactMatch	AF07H:FSSi, AF07H:T011806gt	14,16	28,31	8,10

AF07H:FSSi (FSS-i^3 data) is an exact match to AF07H:T011806gt (Genotyper data)
– thus all allele calls are concordant with AF07H

Uncovering Previous Mistakes During Data Review with Expert System



FOR THE RECORD

John M. Butler,1 Ph.D.; Richard Schoske,1 M.A.; Peter M. Vallone,1 Ph.D.; Janette W. Redman1; and Margaret C. Kline,1 M.S.

Allele Frequencies for 15 Autosomal STR Loci on U.S. Caucasian, African American, and Hispanic Populations*

Failure to click off a low-level peak during data review (allele 8 is not real); not caught by second (thorough) review

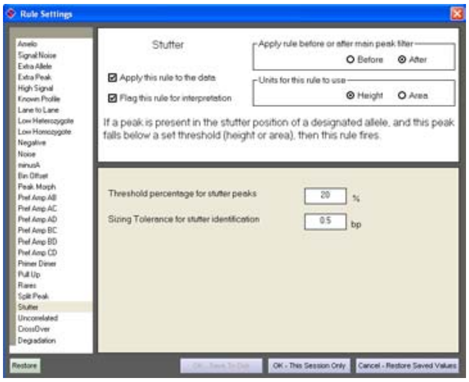
- Data included in Butler *et al.* (2003) *J. Forensic Sci.* 48(4): 908-911
- D13S317 African American allele 8 frequency changes from 0.03295 to 0.03101

Correct call should be 12,12 for D13S317
(Discovered while reviewing FSS-i^3 “discordant” calls)

Types of Error Considered

- Expert system makes a wrong call
- Expert system misses a call

FSS-i^3 Rule Sets and Thresholds



Stutter

☑ Apply this rule to the data

☑ Flag this rule for interpretation

Apply rule before or after main peak filter: ○ Before ● After

Units for this rule to use: ● Height ○ Area

If a peak is present in the stutter position of a designated allele, and this peak falls below a set threshold (height or area), then this rule fires.

Threshold percentage for stutter peaks: 20 %

Sizing Tolerance for stutter identification: 0.5 bp

Batch Summary

262 samples x 16 loci = **4,192 potential allele calls**

Profile Results

- **FP** (full profile) = 3,232 times
- **PP** (partial profile) = 480 times
- **FP-MIX** (full profile with potential mixture) = 688 times
- **PP-MIX** (partial profile with potential mixture) = 208 times
- **NSD** (no signal detected) = 48 times → includes negative controls

Rules Fired

65.2%

- **No rules fired = 2,732 times**
- Signal:Noise = 250 times
- Extra Allele = 269 times
- High Signal = 237 times
- Pref Amp AB = 231 times
- Noise = 157 times
- Peak Morph = 125 times
- Extra Peak = 123 times
- Pull Up = 150 times
- Bin Offset = 45 times
- Degradation = 41 times
- Low Homozygote = 5 times
- Low Heterozygote = 1 time
- Split peak = 1 time

Rules Fired by Locus (1st Set)

Loci	#Total	#0	#1	#2+
TPOX	83	75	6	2
CSF1PO	83	72	9	2
D16S539	83	68	15	0
D13S317	83	67	12	4
AMEL	83	66	17	0
D21S11	83	66	12	5
D5S818	83	65	5	13
FGA	83	64	10	9
D2S1338	83	63	15	5
D18S51	83	62	13	8
D7S820	83	57	25	1
TH01	83	56	22	5
D3S1358	83	52	22	9
D19S433	83	52	13	18
D8S1179	83	40	24	19
VWA	83	32	33	18

This information is output from one of Dave Duewer's programs

All loci had at least one rule fired

D19, D8, and VWA had the most rules fired – most problematic loci in terms of data review

Rules Fired by Locus (2nd Set)

Loci	#Total	#0	#1	#2+
FGA	208	113	24	23
D16S539	208	106	47	7
D13S317	208	106	38	16
D5S818	208	99	19	42
CSF1PO	208	95	51	14
D21S11	208	94	32	34
D7S820	208	92	56	12
D18S51	208	90	41	29
AMEL	208	87	63	10
TPOX	208	87	29	44
D19S433	208	77	32	51
D2S1338	208	72	68	20
TH01	208	60	50	50
D3S1358	208	57	25	78
D8S1179	208	37	39	84
VWA	208	24	33	103

All loci had at least one rule fired

D3, D8, and VWA had the most rules fired – most problematic loci in terms of data review for this set

Rules Fired by Sample

loci where a rule was fired

Sample	#Total	#0	#1	#2+
C18H	16	16	0	0
C30H	16	16	0	0
AF11H	16	15	1	0
AF28H	16	15	1	0
AF93C	16	15	1	0
C11H	16	15	1	0
C19H	16	15	1	0
AF98C	16	15	0	1
AF06H	16	14	2	0
AF97C	16	14	2	0
C04H	16	14	2	0
C28H	16	14	2	0
C15H	16	14	1	1
C23H	16	14	1	1
AF21H	16	14	0	2

(not all samples shown…)

Only two samples had no rules fired

Provides a form of quality checks to the data examined

This information is output from one of Dave Duewer's programs

Concordance Evaluation

- Identifiler data collected on ABI 3100; 262 samples processed in multiple 96-well plates with several allelic ladders
- Typed with GeneScan/Genotyper and GeneMapper ID (MCK)
- Same data processed through GeneMapperID/FSS-i^3 (JMB, BH, AD)
- Results from **245 samples compared** (removed pos. & neg.):
  - 136 samples matched with no data review
  - 109 pairs exhibited a mismatch with **unedited** FSS-i^3 results (including partial profiles)
- **Examination of mismatches to determine which rules fired and if user would be able to make correct calls following editing**

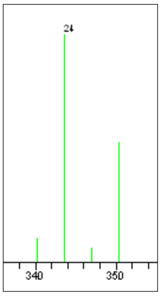
STR_MatchSamples.xls Output Under "Best Match" Showing Several Discordant Calls

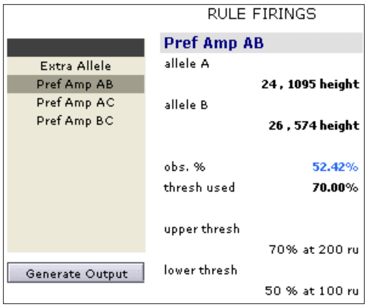
	Example 1 **"Loss" of an Allele**	**Example 2** **"Gain" of an Allele**	**Example 3** **"Wrong Call"**
FSS-i^3 Call (unedited)	D2S1338 ?,24	vWA 16,18	FGA ?,51.2
Genotyper Call	D2S1338 24,26	vWA 18,18	FGA 22,26

Example 1

Unedited FSS-i^3 Data:

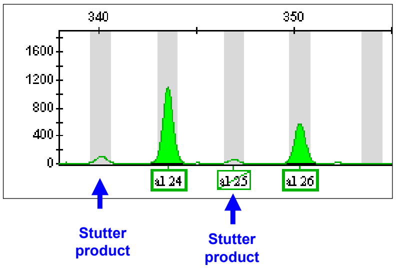
D2S1338 Allele 26 "Loss"





Pref Amp AB
allele A — **24 , 1095 height**
allele B — **26 , 574 height**
obs. % — 52.42%
thresh used — **70.00%**
upper thresh — 70% at 200 ru
lower thresh — 50 % at 100 ru

GeneMapper ID View of D2S1338 Allele 26 "Loss"



Heterozygote peak imbalance

574/1095 =
52.4%

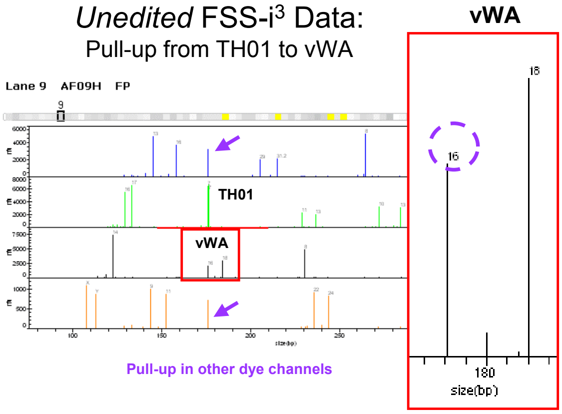
Would re-amp and re-run but because of stutter product allele 26 was not ruled out

---

Example 2

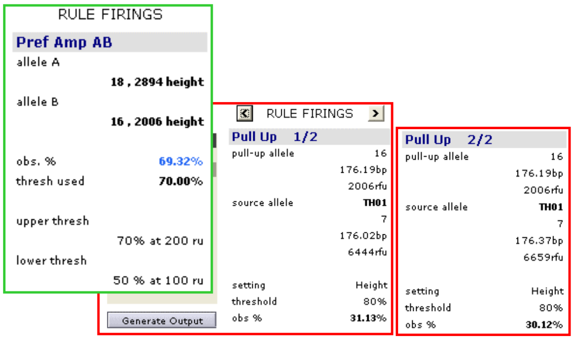
Unedited FSS-i^3 Data:

Pull-up from TH01 to vWA



Pull-up in other dye channels

FSS-i^3 Rules Fired for vWA Locus

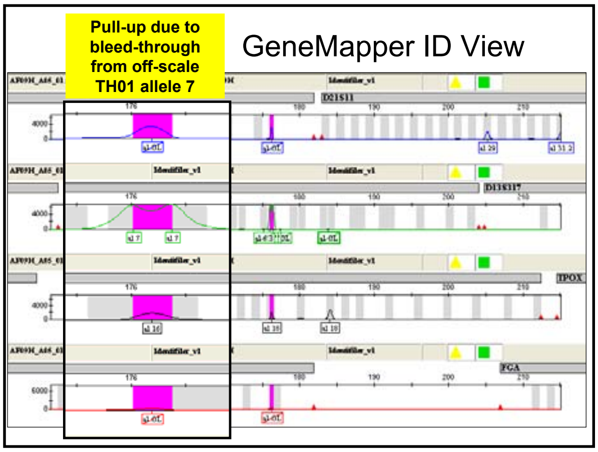


Pref Amp AB
allele A — **18 , 2894 height**
allele B — **16 , 2006 height**
obs. % — 69.32%
thresh used — **70.00%**
upper thresh — 70% at 200 ru
lower thresh — 50 % at 100 ru

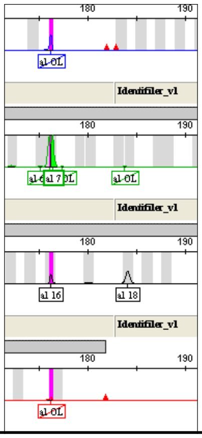
Pull Up 1/2
pull-up allele — 16, 176.19bp, 2006rfu
source allele — **TH01**, 7, 176.02bp, 6444rfu
setting — Height
threshold — 80%
obs % — **31.13%**

Pull Up 2/2
pull-up allele — 16, 176.19bp, 2006rfu
source allele — **TH01**, 7, 176.37bp, 6659rfu
setting — Height
threshold — 80%
obs % — **30.12%**

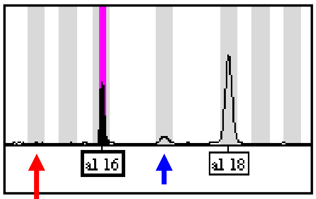
GeneMapper ID View



Pull-up due to bleed-through from off-scale TH01 allele 7



Pull-up due to bleed-through from off-scale TH01 allele 7



No stutter product

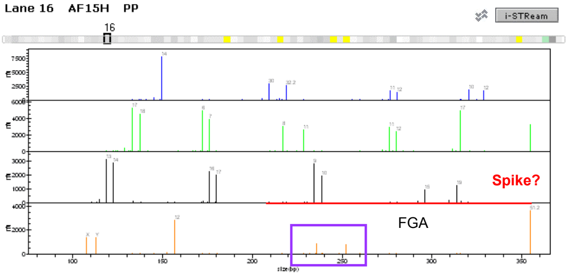
Stutter product

Correct vWA Type = 18,18

Example 3

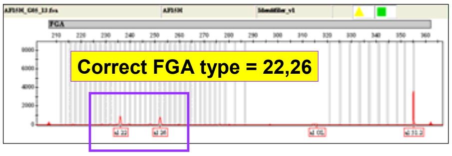
Unedited FSS-i^3 Data:

Spike in FGA High-Molecular Weight Region

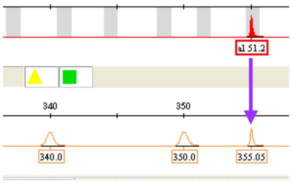


Correct FGA type = 22,26

GeneMapper ID View of FGA “51.2” Spike



Correct FGA type = 22,26



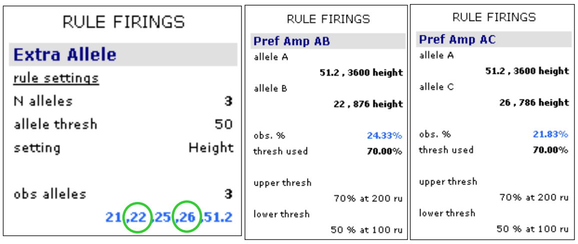
Pull-up due to bleed-through between dye channels

Note the difference in peak shape in the LIZ dye channel relative to the size standard 340 and 350 bp peaks

---

FSS-i^3 Rules Fired

Spike in FGA High-Molecular Weight Region



RULE FIRINGS

Extra Allele

rule settings

N alleles 3

allele thresh 50

setting Height

obs alleles 3

21,22,25,26,51.2

RULE FIRINGS

Pref Amp AB

allele A 51.2 , 3600 height

allele B 22 , 876 height

obs. % 24.33%

thresh used 70.00%

upper thresh 70% at 200 ru

lower thresh 50 % at 100 ru

RULE FIRINGS

Pref Amp AC

allele A 51.2 , 3600 height

allele C 26 , 786 height

obs. % 21.83%

thresh used 70.00%

upper thresh 70% at 200 ru

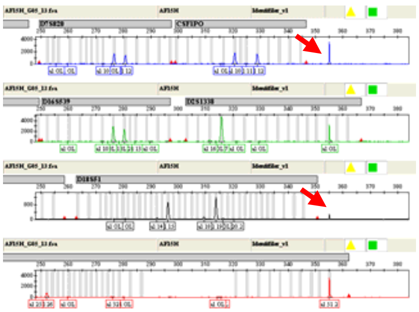
lower thresh 50 % at 100 ru

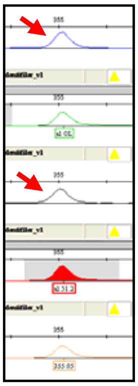
No pull up or peak morph. rule fired!

— probably due to GeneMapperID data extraction problem?

Spike Observed in All Colors

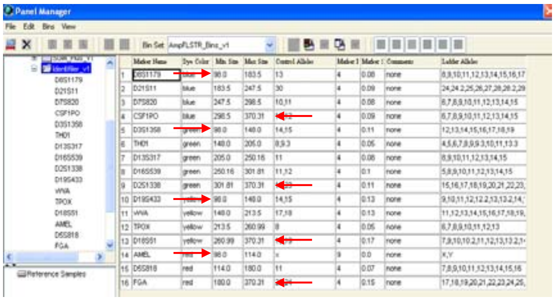
but Not Saved into GeneMapper ID Table





Since the spike was not in allele range for blue and black dye channels it was not included in the peak text file generated for use in FSS-i^3

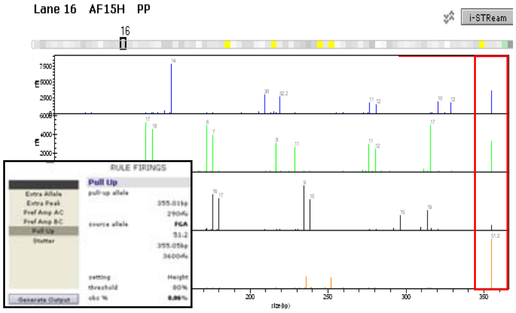
Change GeneMapper Bin Settings



	Marker Name	Dye Color	Min Size	Max Size	Control Alleles	Marker Repeat	Marker Spec	Comments	Ladder Alleles
1	D8S1179	blue	98.0	183.5	13	4	0.08	none	8,9,10,11,12,13,14,15,16,17
2	D21S11	blue	183.5	247.5	30	4	0.09	none	24,24.2,25,26,27,28,28.2,29
3	D7S820	blue	247.5	298.5	10,11	4	0.08	none	6,7,8,9,10,11,12,13,14,15
4	CSF1PO	blue	298.5	370.31	10,12	4	0.09	none	6,7,8,9,10,11,12,13,14,15
5	D3S1358	green	98.0	140.0	14,15	4	0.11	none	12,13,14,15,16,17,18,19
6	TH01	green	140.0	205.0	8,9.3	4	0.05	none	4,5,6,7,8,9,9.3,10,11,13.3
7	D13S317	green	205.0	250.16	11	4	0.08	none	8,9,10,11,12,13,14,15
8	D16S539	green	250.16	301.81	11,12	4	0.1	none	5,8,9,10,11,12,13,14,15
9	D2S1338	green	301.81	370.31	19,23	4	0.11	none	15,16,17,18,19,20,21,22,23,
10	D19S433	yellow	98.0	140.0	14,15	4	0.13	none	9,10,11,12,12.2,13,13.2,14,
11	vWA	yellow	140.0	213.5	17,18	4	0.13	none	11,12,13,14,15,16,17,18,19,
12	TPOX	yellow	213.5	260.99	8	4	0.05	none	6,7,8,9,10,11,12,13
13	D18S51	yellow	260.99	370.31	15,19	4	0.17	none	7,9,10,10.2,11,12,13,13.2,1
14	AMEL	red	98.0	114.0	X	9	0.0	none	X,Y
15	D5S818	red	114.0	180.0	11	4	0.07	none	7,8,9,10,11,12,13,14,15,16
16	FGA	red	180.0	370.31	23,24	4	0.15	none	17,18,19,20,21,22,23,24,25,

Min/Max Sizes were adjusted to be the same for all dye channels, and marker bin settings were lined up so there were no gaps within a dye channel

Spike displayed in ALL dye channels

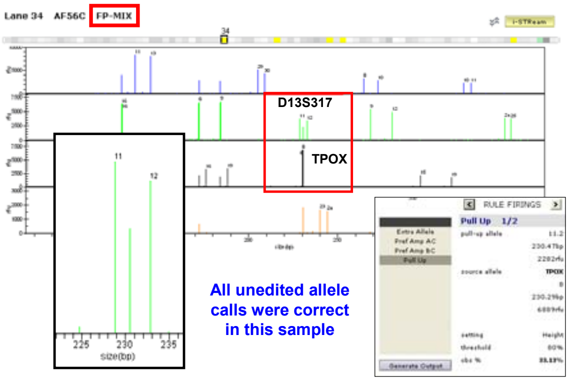


The spike is still being called a 51.2, and pull up is only being called in the yellow channel

STR_MatchSamples.xls Output Under "Best Match" Showing Several Discordant Calls

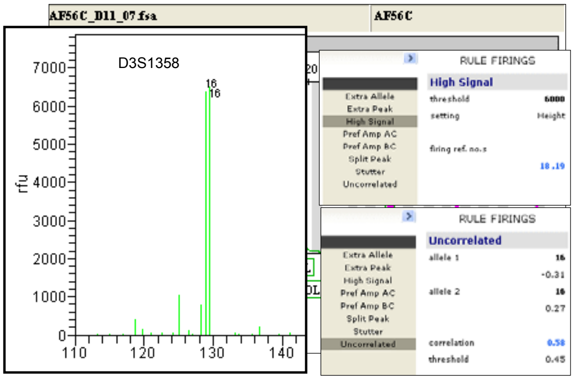
	Example 1 "Loss" of an Allele	**Example 2** "Gain" of an Allele	**Example 3** "Wrong Call"
FSS-i³ Call (unedited)	D2S1338 ?,24	vWA 16,18	FGA ?,51.2
Genotyper Call	D2S1338 24,26	vWA 18,18	FGA 22,26
	Correct call is 24,24 (using 70% threshold)	Call would likely be edited to 18,18 (due to pull-up rule firing)	Pull-up and peak morph. rules did not fire but an experienced analyst would probably make the correct call during data review
FSS-i³ Call (edited) - after checking rule firings	D2S1338 24,26 ?	vWA 18,18	

Extreme Pull-Up Due to Off-Scale Peaks

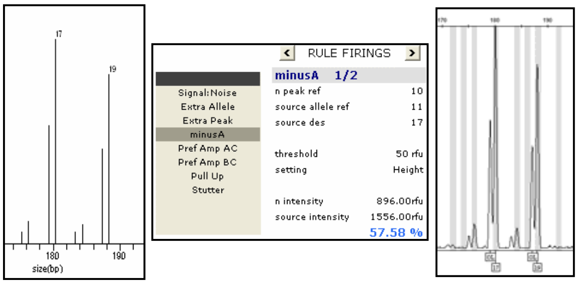


All unedited allele calls were correct in this sample

Off-Scale D3S1358 Peak



Minus A at vWA



FSS-i³

GeneMapper ID

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

If No Rules Fired, Were There Any Mistakes?

- 2,732 genotypes made with “no rules fired” (4,192 possible types across 262 samples) – still must click through most samples
- Each genotype was carefully re-reviewed manually
- **No discrepancies (discordance) were noted in calls based on rules set**
- **This observation provides confidence that when no rules are fired, data quality is acceptable in the data sets reviewed thus far…**

Reviewing a Large Data Set

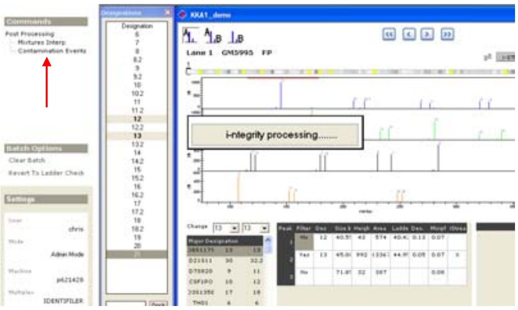
Nice Features

- Rapid check of all allelic ladders and generation of composite allelic ladders
- Rapid processing of data

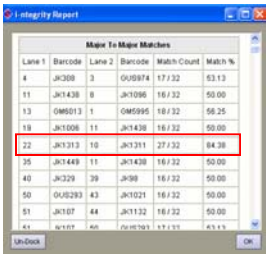
Cumbersome Features

- Having to click through every sample in order to review rule firings

i-ntegrity Contamination Evaluation



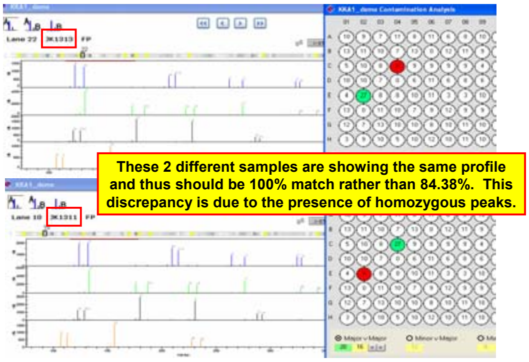
i-ntegrity Contamination Analysis



Major To Major Matches					
Lane 1	Barcode	Lane 2	Barcode	Match Count	Match %
4	JK308	3	OUS974	17 / 32	53.13
11	JK1438	8	JK1096	16 / 32	50.00
13	GM6013	1	GM5995	18 / 32	56.25
19	JK1006	11	JK1438	16 / 32	50.00
22	JK1313	10	JK1311	27 / 32	84.38
35	JK1449	11	JK1438	16 / 32	50.00
40	JK329	39	JK98	16 / 32	50.00
50	OUS293	43	JK1021	16 / 32	50.00
51	JK107	44	JK1132	16 / 32	50.00

These 2 samples are showing a high match % of 84.38, so we want to examine these samples further

i-ntegrity Contamination Analysis



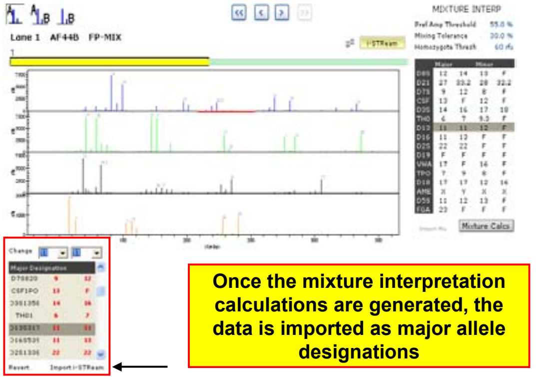
These 2 different samples are showing the same profile and thus should be 100% match rather than 84.38%. This discrepancy is due to the presence of homozygous peaks.

Issues with i-ntegrity

- The i-ntegrity contamination analysis is not in a 96 well plate format. Instead, it is in the 16 capillary format of the 3100/3130xl making it difficult to locate samples.
- A 100% match is only possible if the two samples being compared are heterozygous for all loci. If the samples are exactly the same and are homozygous for some loci, it will not register as a 100% match which can be misleading.

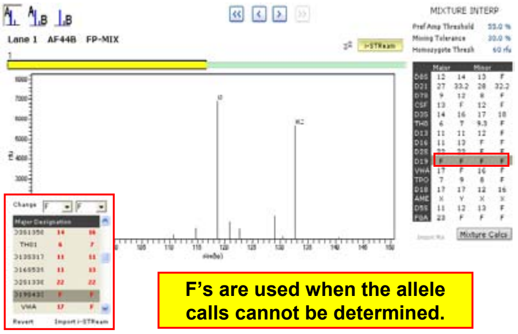
i-STReam Mixture Analysis

Mixture Interpretation Results



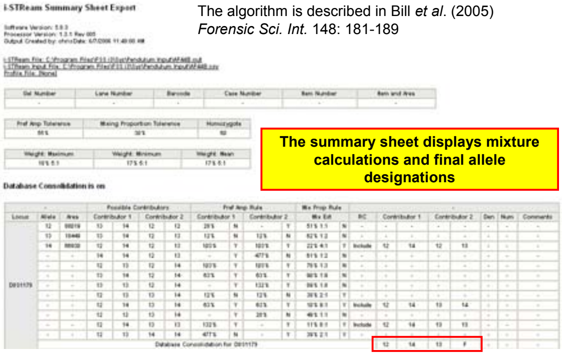
Once the mixture interpretation calculations are generated, the data is imported as major allele designations

Major Allele Designations



F's are used when the allele calls cannot be determined.

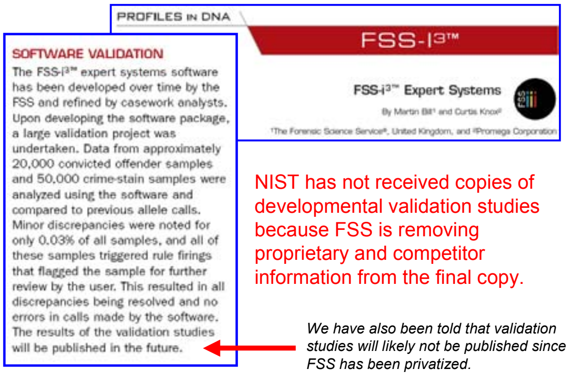
i-STReam Summary Sheet



The algorithm is described in Bill *et al.* (2005)
Forensic Sci. Int. 148: 181-189

The summary sheet displays mixture calculations and final allele designations

Developmental Validation Studies?



PROFILES IN DNA

FSS-i3™

FSS-i3™ Expert Systems

By Martin Bill¹ and Curtis Knox²

¹The Forensic Science Service®, United Kingdom, and ²Promega Corporation

SOFTWARE VALIDATION

The FSS-i3™ expert systems software has been developed over time by the FSS and refined by casework analysts. Upon developing the software package, a large validation project was undertaken. Data from approximately 20,000 convicted offender samples and 50,000 crime-stain samples were analyzed using the software and compared to previous allele calls. Minor discrepancies were noted for only 0.03% of all samples, and all of these samples triggered rule firings that flagged the sample for further review by the user. This resulted in all discrepancies being resolved and no errors in calls made by the software. The results of the validation studies will be published in the future.

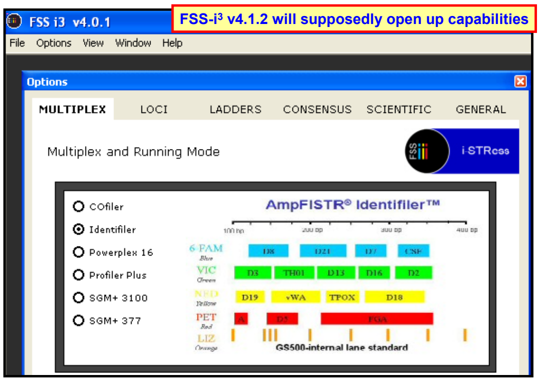
NIST has not received copies of developmental validation studies because FSS is removing proprietary and competitor information from the final copy.

We have also been told that validation studies will likely not be published since FSS has been privatized.

What FSS-i^3 cannot do…

- Process the following kits:
  - Profiler STR kit
  - Yfiler or PowerPlex Y kits (e.g., we have >1,200 Yfiler profiles available at NIST)
  - Custom assays (e.g., miniSTRs)
- Input/Output format issues:
  - Once data input and output formats have been created, cannot pull up formats to view and modify

Currently Supported STR Kits



FSS-i^3 v4.1.2 will supposedly open up capabilities

○ COfiler
● Identifiler
○ Powerplex 16
○ Profiler Plus
○ SGM+ 3100
○ SGM+ 377

Thoughts Regarding FSS-i^3 Software

- There is a learning curve with the software
- Much faster to process data but **full data review can be lengthy**
- Must examine rule firings—cannot just accept unedited data

What We Would Like to Have Improved

Suggestions for next update:

- Detailed User Manual
- Capability of processing more kits and custom assays
- Modify sample position layout in i-ntegrity to be a 96-well format
- Permit allelic ladder to be in the A01 position
- View data input/output formats and edit them
- Save changes during session
- System shuts down if 50/50 mixtures are present
  - Accommodate for this problem in i-STReam mixture module even if allele calls cannot be made

The newest version 4.1.2 of the software accommodates for all of these suggestions and is coming soon…

Possibility of NIST Standard Data Set

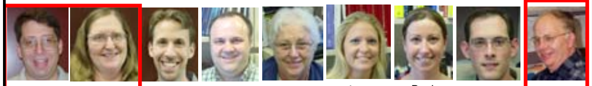
- Is there any interest?
- Data set of .fsa files could be made available for download from STRBase as WinZip files
- Could be used for verifying allele calls with new allele calling software or upgrades to existing expert systems running in your lab

Future Plans

- MIX05 Interlab mixture data sets have been run and data is currently under review to evaluate i-STReam module
- More data sets are available at NIST and will be processed with FSS-i^3 for comparison purposes
- Release additional software tools on STRBase (Dave Duewer programs)
- Publish recommendations on approaches for validation of expert system software

Acknowledgments

Funding from interagency agreement 2003-IJ-R-029 between NIJ and the NIST Office of Law Enforcement Standards



John Butler
Margaret Kline
Pete Vallone
Mike Coble
Jan Redman
Amy Decker
Becky Hill
Chris DeAngelis
Dave Duewer

Software Installation at NIST:
Bob McLaren

http://www.promega.com/profiles/901/ProfilesInDNA_901_16.pdf

Training at Promega:
Bob McLaren
Kim Huston
Curtis Knox