Multiplexes, MiniSTRs, and Mixtures: Uses for the FSS-i³ Software at NIST

Becky Hill

Amy E. Decker, Angie M. Dolph, Margaret C. Kline and
John M. Butler

National Institute of Standards and Technology

**Promega Summer Technology Tour – August 2007
Cupertino, CA**

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**

Outline of Topics to Discuss

- Overview of the FSS-i³ software
- How to create a new multiplex in multiplex manager, i.e. MiniFiler
- Work performed at NIST
  - Single source samples
  - Mixture samples
- Conclusions and future plans

Overview of the FSS-i³ Software

Cost to the End User

- **Software**
  - Receive the v4.1.3 upgrade software
  - single copy, single computer **$20,000**

- **Maintenance agreement**
  - **$4,000** per year (20% of total software cost per year, max $15,000)
  - Software upgrades and patches 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
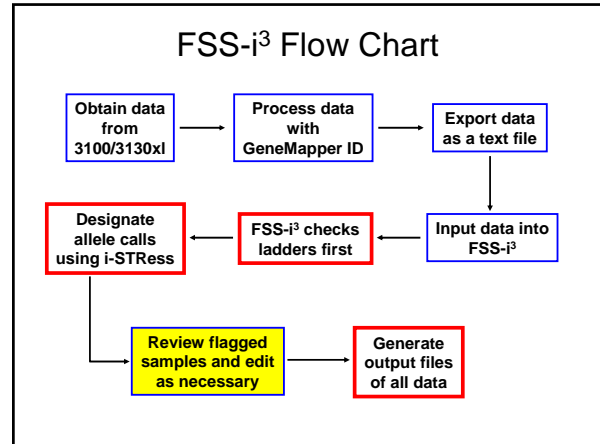
Minimum starting cost of $26,000

Overview of Software Components

integrity
interpretation
innovation

i-STRess   i-STRream   i-ntegrity

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-STRream 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

FSS-i³ Flow Chart

Obtain data from 3100/3130xl → Process data with GeneMapper ID → Export data as a text file

Designate allele calls using i-STRess ← FSS-i³ checks ladders first ← Input data into FSS-i³

Review flagged samples and edit as necessary → Generate output files of all data

Change GeneMapper ID Bin Settings

These bins were developed for FSS-i³ by Promega

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
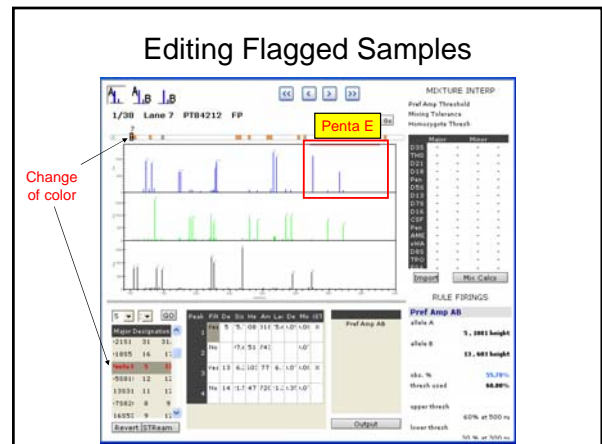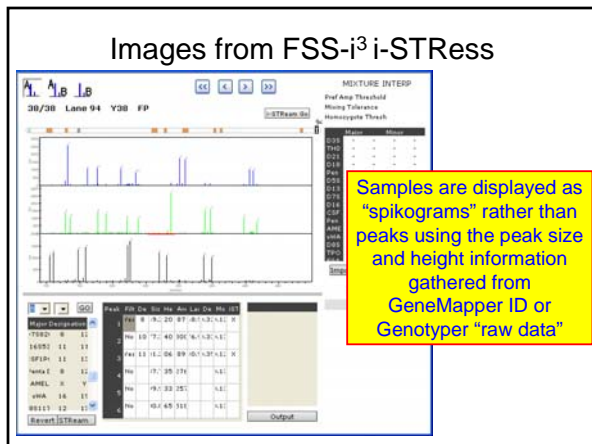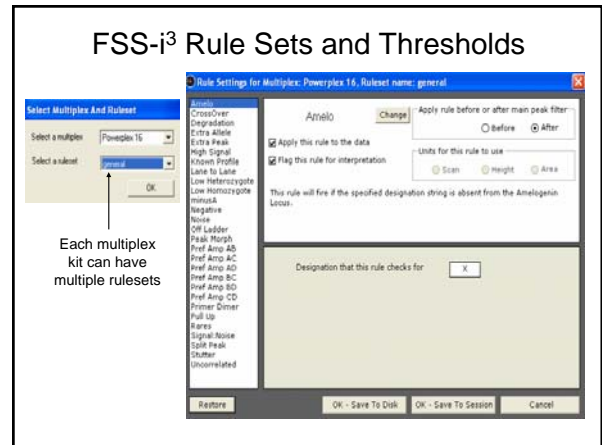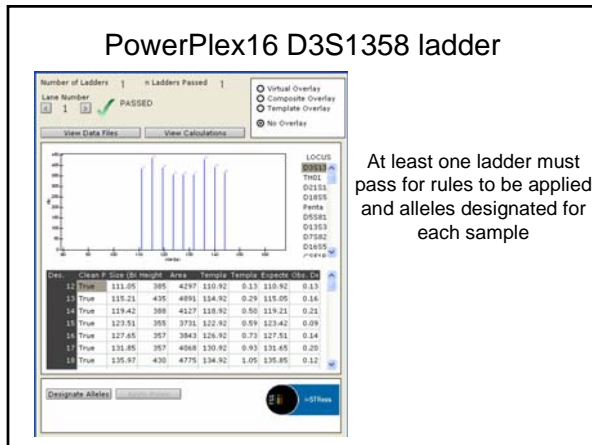
Change GeneMapper ID Bin Settings

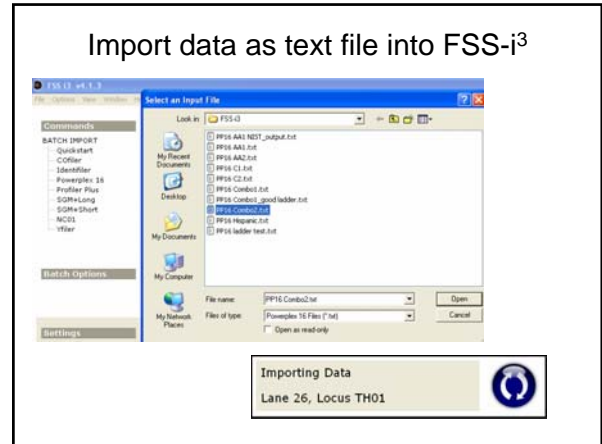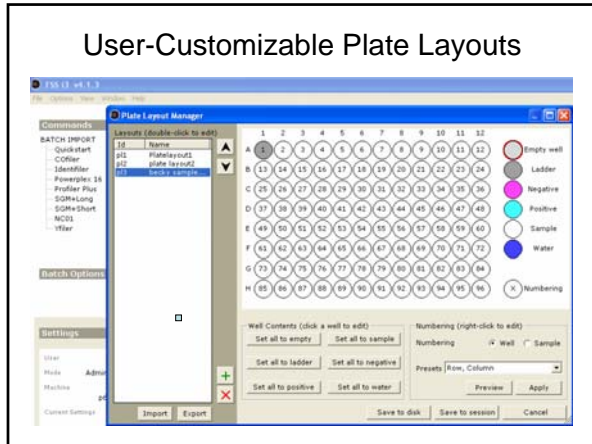**Example Identifiler ladder**

Gap between D3 and TH01 loci
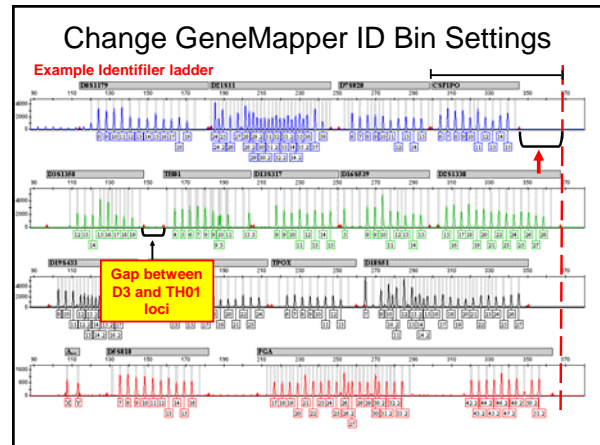
GeneMapper ID Raw Data Sorted

FSS-i³ Input Template

Custom assays not provided

Controls what and how information will be brought into the FSS-i³ program

User-Customizable Plate Layouts

Import data as text file into FSS-i³

PowerPlex16 D3S1358 ladder

At least one ladder must pass for rules to be applied and alleles designated for each sample

FSS-i³ Rule Sets and Thresholds

Each multiplex kit can have multiple rulesets

Images from FSS-i³ i-STRess

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

Editing Flagged Samples

Change of color

Penta E

FSS-i³ Output Template

Controls what and how information will be exported from the FSS-i³ program

Output Files Created by FSS-i³ i-STRess

C:\Program Files\FSSi3 v412d\i-STRess\Files\PP16 Combo2 Output Files

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

Audit Files Created by FSS-i³ i-STRess

C:\Program Files\FSSi3 v412d\i-STRess\Files\PP16 Combo2 Audit Files

FSS-i³ data review can be saved as a "Batch"

How to Create a New Multiplex in Multiplex Manager

Creating new multiplexes with Multiplex Manager

Enter the Name of the New Multiplex

Designate the sequencer

Change parameters in Advanced Ladder Settings

Click Templates button and designate all names of ladders used

Adding new loci and allele designations

Entering first allele in ladder
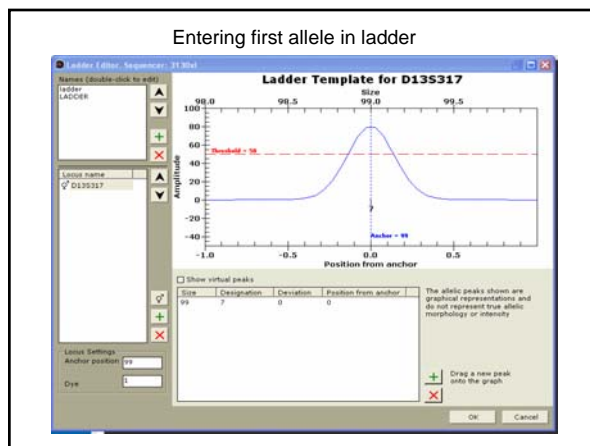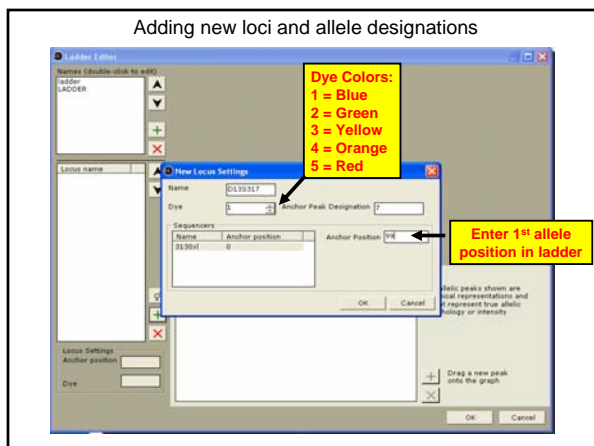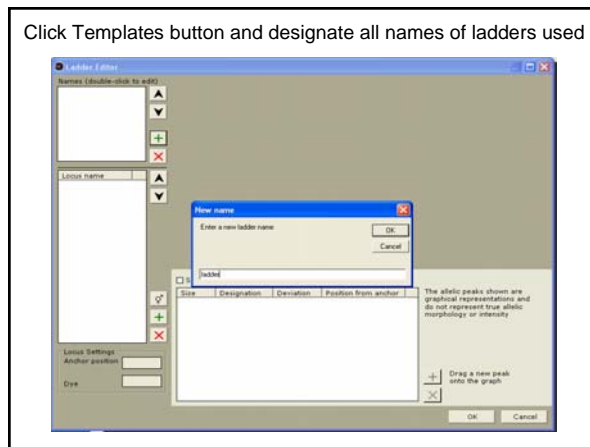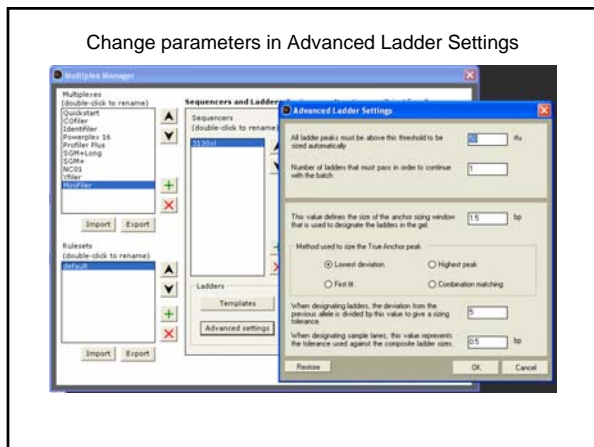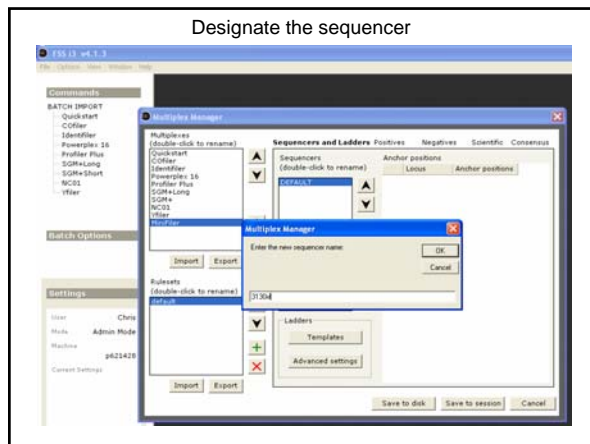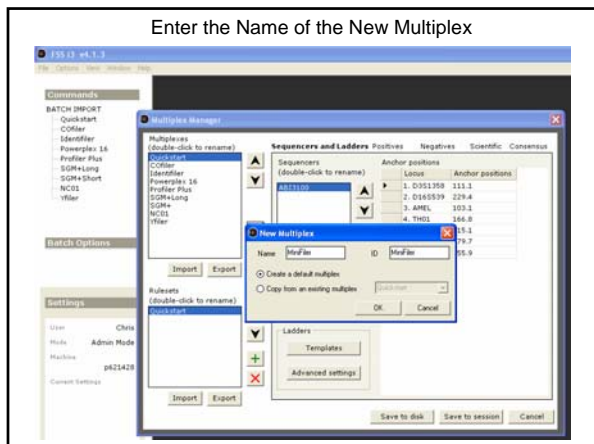
Add subsequent alleles from initial anchor position

Position from anchor depends on locus repeat number (i.e. D13S317 is a 4 bp repeat, so each allele is positioned accordingly

Adding variant alleles
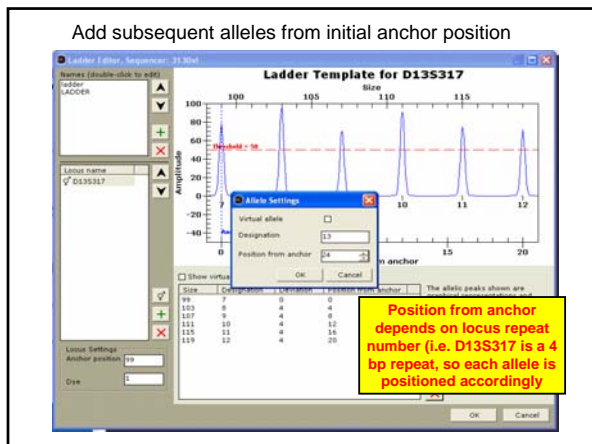
The deviation is different for variant alleles
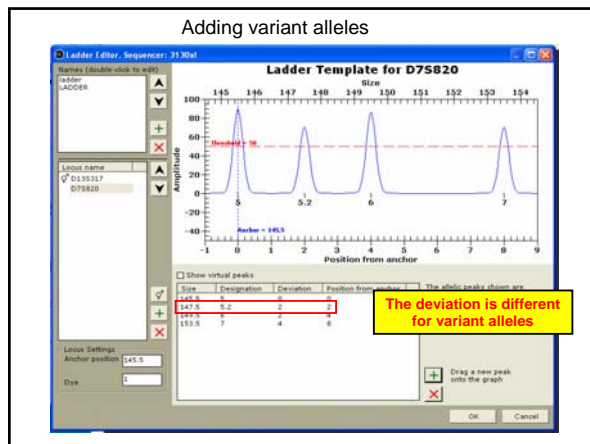
Complete ladder before designation of virtual alleles
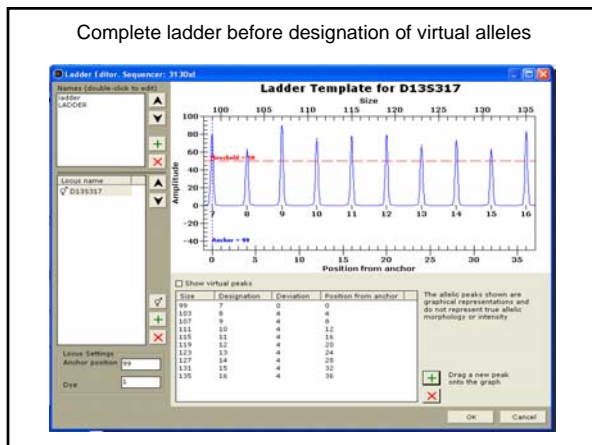
Virtual allele designation, anchor point should be 1st **REAL** allele
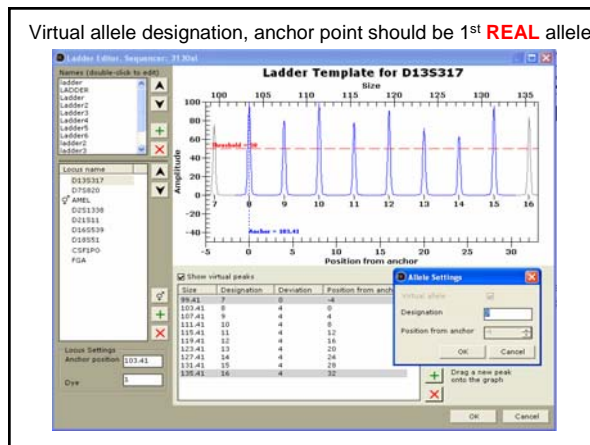
When the multiplex is complete, all anchor positions are displayed
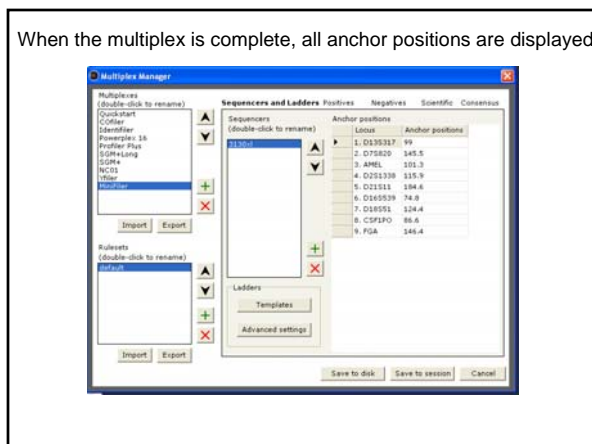
Creation of a new ruleset for the multiplex

Adding positive controls to the multiplex

Adding negative controls to the multiplex

Ladder for MiniFiler D13S317

Spikogram view of MiniFiler data

Ladder editor for Y-filer multiplex

Y-filer DYS456 ladder

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

Spikogram view of Y-filer data
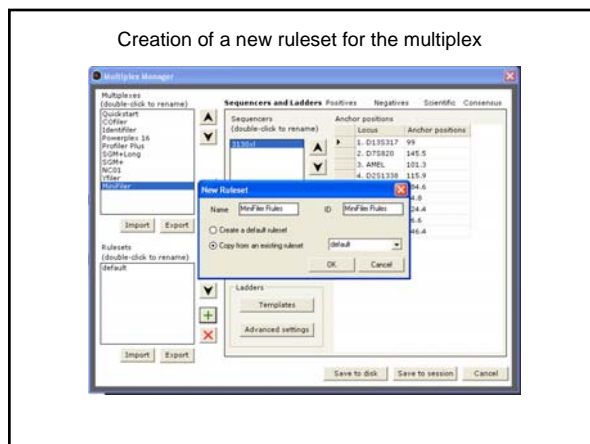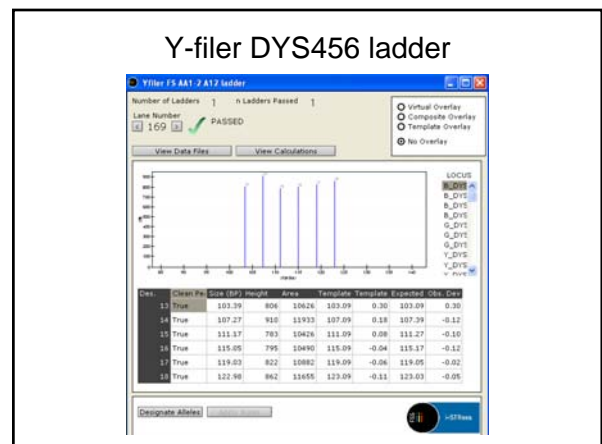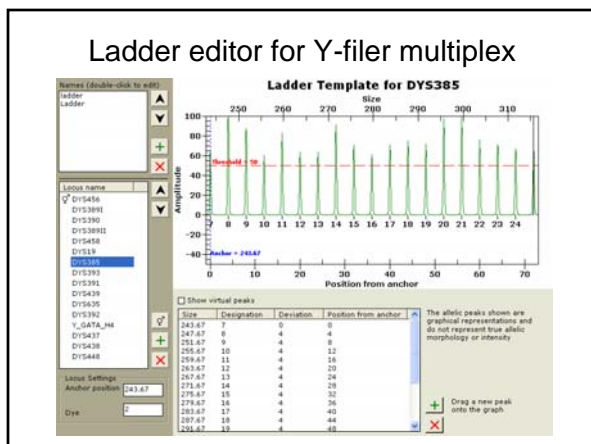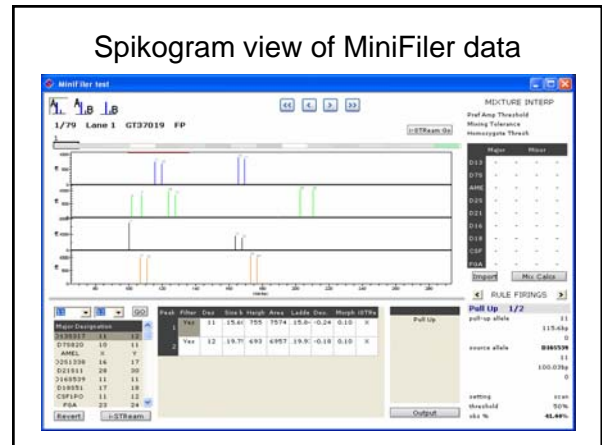
Work Performed at NIST
Single Source Samples

Allele Concordance Studies at NIST with Single Source Samples

- Manual calls
  - with GeneScan/Genotyper v3.7
  - with GeneMapper ID v3.2

- Automated calls with GM/FSS-i³

- Comparison of output with Excel spreadsheets written by Dave Duewer (NIST)

Single Source Samples Examined with i-STRess

- We have previously examined **262 Identifiler** samples with v4.0.1 and **656 PowerPlex16** samples with v4.1.3. Excellent concordance was found and the results can be found in past presentations:

  http://www.cstl.nist.gov/biotech/strbase/pub_pres/NIST_FSSi3_Mar2006.pdf
  http://www.cstl.nist.gov/biotech/strbase/pub_pres/PromegaTechTour_NIST_FSSi3.pdf
  http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2006_FSSi3.pdf

- In this presentation I will present results from **982 MiniFiler** samples run with v4.1.3.

Data Comparison Between Methods

- Dave Duewer (NIST Analytical Chemistry Division) has written several computer programs to convert and compare FSS-i³ data that utilize Excel macros

  - DNA_FSSi3_Convert.xls (converts data format)
  - STR_MatchSamples.xls (compares samples)

- These programs are currently available to the community
  - **http://www.cstl.nist.gov/biotech/strbase/software.htm**

DNA_FSSi3_Convert.xls

First five columns in FSS-i³ output are converted to be like GeneMapper ID allele designation table

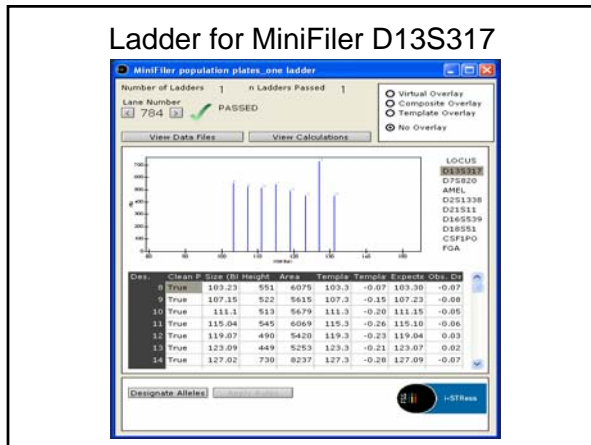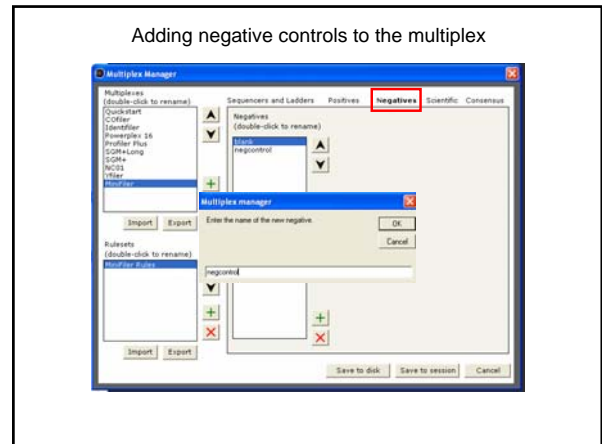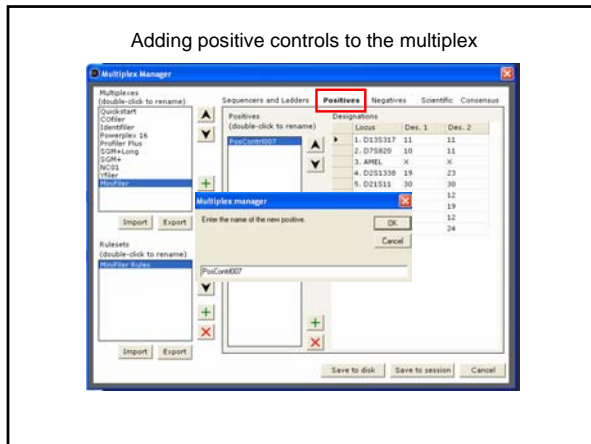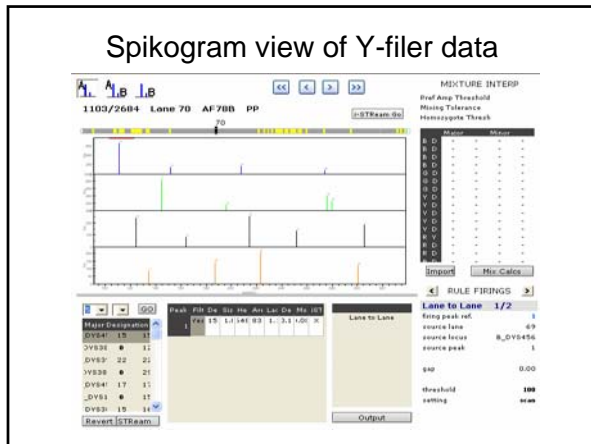| Batch ID	Sample ID	Locus ID	Major Designation 1	Major Designation 2
MiniFiler RAW AA	GT37019	D13S317	11	12
MiniFiler RAW AA	GT37019	D7S820	10	11
MiniFiler RAW AA	GT37019	AMEL	X	Y
MiniFiler RAW AA	GT37019	D2S1338	16	
MiniFiler RAW AA	GT37019	D21S11	28	
MiniFiler RAW AA	GT37019	D16S539	11	
MiniFiler RAW AA	GT37019	D18S51	17	
MiniFiler RAW AA	GT37019	CSF1PO	11	12
MiniFiler RAW AA	GT37019	FGA	23	24

Each row is an individual locus

Data Transformation

SampleCode	D13S317	D13S317	D7S820	D7S820	AMEL	AMEL	D2S1338	D2S1338	D21S11	D21S11	D16S539	D16S539	D18S51	D18S51	CSF1PO	CSF1PO	FGA	FGA
GT37019	11	12	10	11	X	Y	16	17	28	30	11	11	17	18	11	12	23	24
GT37020	11	13	9	11	X	Y	19	26	28	28	11	12	16	16				
GT37026	12	12	9	10	X	Y	17	21	28	28	11	12	12	17				
GT37027	12	13	8	10	X	Y	17	20	29	31	9	12	15	17				
GT37032	11	12	8	8	X	Y	20	22	29	31	9	11	17	19				

Each row is an individual sample

STR_MatchSamples.xls

Microsoft Excel - STR_MatchSamples.xls

File Edit View Insert Format

	#Samples	#ExactMatch	#Unmatched
	95	67	28

A20

	Status	Date	Sheet	Contains
Clear All	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'

First Data

Add Data

Two or more data sets can be compared to one another

Exact Match

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

Best Match

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

Type	Sample Description	Penta_D	Penta_E	TH01	TPOX	vWA
Unmatched	GT37019:all data	2.2,11	5,13	6,7	8,9	17
Unmatched	GT37019:all PP16 GM samples	2.2,11	5,13	6,7	8,9	14,17
ExactMatch	BC11352:all data, BC11352:all PP16 GM samples	10,11	7,12	6,9.3	8	14,17
ExactMatch	GA05070:all data, GA05070:all PP16 GM samples	13,14	7,17	7,9	8,12	14,19
ExactMatch	GA05071:all data, GA05071:all PP16 GM samples	10,11	11,12	7,9.3	8,11	16,17
ExactMatch	GC03394:all data, GC03394:all PP16 GM samples	10,11	12,15	6,7	8	17,18

•Unmatched sample type flags discordant calls
•ExactMatch sample type indicates full concordance between FSS-i³ and GeneMapper ID samples

---

Concordance Evaluation

- MiniFiler collected on ABI 3130xl; 982 samples processed in GeneMapper ID and FSS-i³

- Typed manually with GeneMapper ID
- Same data processed through GeneMapper ID/FSS-i³
- When rules were fired, profiles were reviewed

- Results from **982 samples** compared with STR_MatchSamples.xls

- **Examination of mismatches to determine which rules were fired and if user would be able to make correct calls following editing: All calls were concordant after review**

Example 1- Microvariant
Unedited FSS-i³ Data:
D21S11 Allele 33.1

RULE FIRINGS

Signal Noise   **Extra Peak**
Extra Peak   rule settings
                n peaks   1
                peak thresh   200
                setting   height
                no. peaks & refs   1

No virtual bin for this microvariant was entered into software when MiniFiler was created therefore this allele was not being called in FSSi³

Example 2- MiniFiler Issue
Unedited FSS-i³ Data:
D16S539 Alleles not called by software

RULE FIRINGS

Pref Amp AB
Pref Amp AB   allele A
Pull Up   12 , 1860 height
               allele B
                             11 , 295 height

               obs. %:   15.86%
               thresh used   50.00%

               upper thresh
                             50% at 250 ru
               lower thresh
                             45 % at 150 ru

               Output

Pref Amp Rule firing due to peak imbalance.

Concordance Studies Reveal Potential Primer Binding Site Mutations with Different Primer Sets

Conventional PCR primer   miniSTR primer

                      *

          STR repeat region

                        miniSTR   Conventional
                        primer   PCR primer

Identifier

Carolyn R. Hill,[1] M.S.; Margaret C. Kline,[1] M.S.; Julio J. Mulero,[2] Ph.D.; Robert E. Lagacé,[2] B.A.; Chien-Wei Chang,[2] Ph.D.; Lori K. Hennessy,[2] Ph.D.; and John M. Butler,[1] Ph.D.

Concordance Study Between the AmpFℓSTR® MiniFiler™ PCR Amplification Kit and Conventional STR Typing Kits*

[1]National Institute of Standards and Technology, Biochemical Science Division, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899.

[2]Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404.

MiniFiler (beta-test)

If No Rules Fired, Were There Any Mistakes?

- Each genotype was carefully re-reviewed with STR_MatchSamples.xls

- **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…**
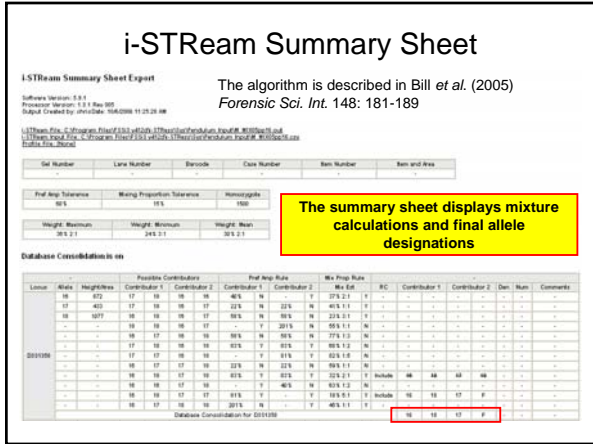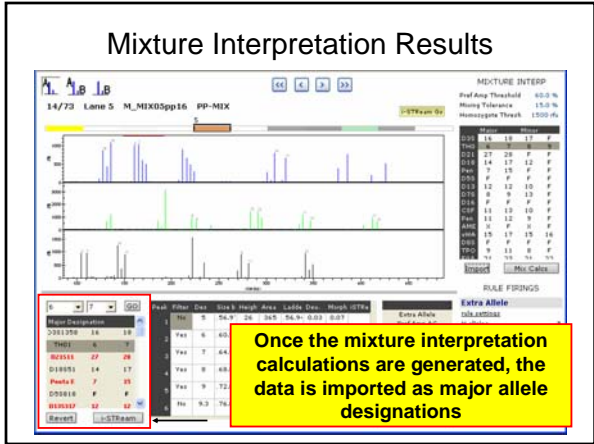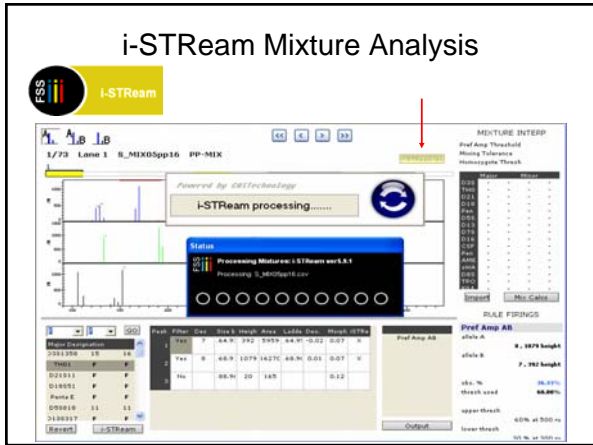
Reviewing a Large Data Set

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

Work Performed at NIST
Mixture Samples

i-STReam Mixture Analysis

Mixture Interpretation Results

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

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

Experiment 1 – MIX05 Data
Mixture Deconvolution
Experimental Design

- Several STR kits were used in the study:
  - SGM+, Profiler Plus, Identifiler, COfiler, Powerplex 16

- 3100 Genetic Analyzer was used to process the samples

- Data was previously collected and profiles analyzed in GMID v3.2

- Mixture deconvolution:
  - FSS-i³ i-STReam, stand-alone version

i-STReam Results

Case 1
3 female : 1 male

Case 2
1 female : 3 male

Case 3
1 female : 1 male

Case 4
7 female : 1 male

i-STReam

Stand-alone version

Case #	Genotypes Called	Accuracy
1	74	100
2	82	100
3	69	83
4	68	98
Overall	73	95

Experiment 2 – Replicates and Ratios
Mixture Deconvolution
Experimental Design

- Identifiler, COfiler, Profiler Plus kits were used

- 1:2, 1:3, 1:5, and 1:8 mixture ratios were prepared

- 6-7 amplification replicates for each mix ratio
  - To determine PCR variation

- How does i-STReam handle this variation?
  - Different results for the same mixture?
  - Incorrect calls?

Experimental Design: Plate Layouts

Identifiler / COfiler 1:2 and 1:5

Profiler Plus 1:2 and 1:5

Identifiler 1:3 and 1:8

COfiler and Profiler Plus 1:3 and 1:8

1:2 and 1:5 all male genotypes

1:3 and 1:8 one female and one male

Profiler Plus opposite major and minor contributors

Total i-STReam Results

Total # Alleles	% 4-Allele	% 3-Allele	% 2-Allele	% 1-Allele
4080	25	33	38	4

i-STReam Results According to Mixture Ratio

- 1:2 ratio worst results
56% Correct Allele Calls

- 1:3 ratio best results
78% Correct Allele Calls

- Drop-out observed in 1:5 and 1:8 Ratios

i-STReam Results According to STR Kit

i-STReam Replicate and Ratio Results According STR Kit

Identifiler				
Total # Alleles	% 4 Allele	% 3 Allele	% 2 Allele	% 1 Allele
2352	29	33	33	5

Profiler Plus				
Total # Alleles	% 4 Allele	% 3 Allele	% 2 Allele	% 1 Allele
1000	35	25	35	5

COfiler				
Total # Alleles	% 4 Allele	% 3 Allele	% 2 Allele	% 1 Allele
728	0	43	57	0
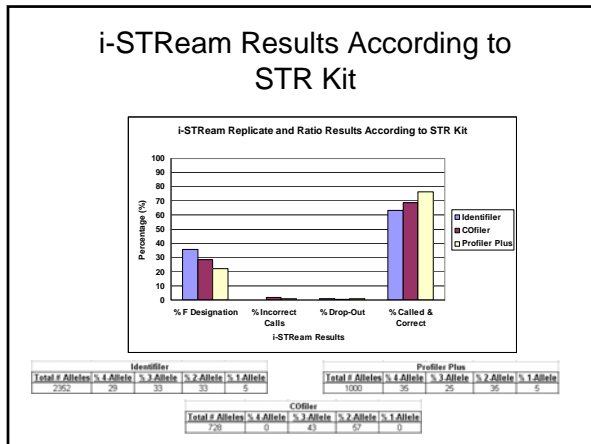
Some i-STReam Observations…

- GeneMapper ID minus A and stutter filters set at zero to allow all alleles into FSS-i³
  - Some minor alleles filtered out as stutter and not called

- Some incorrect calls
  - Incorrect calls can be explained by variation in peak height ratios
  - 26 / 4080 alleles (0.64%)

- Very conservative
  - F designations allow the program to not make a call

Conclusions and Future Plans

In Summary

- FSS-i³ has the capability to create new multiplex kits (Y-filer, PowerPlex Y, MiniFiler and custom assays)

- Dave Duewer software programs are currently available on STRBase:
  http://www.cstl.nist.gov/biotech/strbase/software.htm

- A total of **2162** profiles have been analyzed using the FSS-i³ software with concordance checks performed. Full concordance has been achieved after careful review.

- In general, FSS-i³ i-STReam is conservative in its mixture deconvolution; however, only **26** out of **4080** allele calls were called incorrectly (**0.64%**).

Future Plans

- We plan to explore i-STReam capabilities further

- We will run more data sets that are available at NIST

- Publish recommendations on approaches for validation of expert system software

---

Acknowledgments

John Butler   Margaret Kline   Pete Vallone   Jan Redman   Amy Decker   Becky Hill   Dave Duewer

*Special thanks to **Angie Dolph**, our summer intern from Marshall U. for the mixture study data

Software Installation at NIST:
Bob McLaren

Training at Promega:
Bob McLaren
Kim Huston
Curtis Knox

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