NIST Experience with FSS-i3 Software

Carolyn R. "Becky" Hill,
Margaret C. Kline and John M. Butler
National Institute of Standards and Technology

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
  – Unclear whether or not 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

Packaging for FSS i-Cubed Software

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

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

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

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
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-i3 Rule Sets
- **Day 3:** Settings folder, output templates, i-integrity 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-i3 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-integrity** 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.

Features of FSS-i3 listed on Promega website

- 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-i3 Software

- Obtain data from 3100/3130xl
- Process data with GeneMapper ID
- Export data as a text file
- Input data into FSS-i3
- Designate allele calls using i-STRess
- FSS-i3 checks ladders first
- Generate output files of all data
- Review flagged samples and edit as necessary

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
FSS-i³ Input Template

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

GeneMapper ID Raw Data Sorted

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

FSS-i³ Output Templates

Output Files Created by FSS-i³ i-STRess

C:\Program Files\FSS i³\Files\FS_Identifiers_84 Output Files

Output Files Created by FSS-i³ i-STRess

C:\Program Files\FSS i³\Files\FS_Identifiers_84 Audit Files

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

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Images from FSS-i3

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

Identifier D2S1338 Allelic Ladder

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 AFDL QC sample checks

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

Single Source Samples Examined with i-STRess

- Ran 84 Identifiler samples (father-son samples) with GM/FSS-i3 and compared to GeneScan/Genotyper and GeneMapper ID results
- Ran 864 Identifiler samples (700 reported NIST U.S. population samples) with GM/FSS-i3 and compared to GeneScan/Genotyper results (see http://www.cstl.nist.gov/biotech/strbase/NISTpubdata/JFS2003Iresults.xls)

Over 1,200 unique samples will be examined eventually

Work Performed at NIST

- 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)
Issues with Review of Previous Data

- Need a rapid way to compare allele calls for concordance purposes
  - Allele calls from Genotyper are in different format from FSS-i3 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-i3 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

DNA_FSSi3_Convert.xls
First five columns in FSS-i3 output are converted to be like Genotyper allele designation table

<table>
<thead>
<tr>
<th>Match Name</th>
<th>Sample ID</th>
<th>Locus ID</th>
<th>Major Designation 1</th>
<th>Major Designation 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>D8S179</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>D21S11</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>D7S820</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>CSF1PO</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>D5S1358</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>TH01</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>D13S317</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>ID_FSpairs_1</td>
<td>AF01C</td>
<td>D18S539</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

Each row is an individual sample

Data Transformation

Each row is an individual locus

STR_MatchSamples.xls
Two or more data sets can be compared to one another

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

<table>
<thead>
<tr>
<th>Sample</th>
<th>CB5L179</th>
<th>CB5L179</th>
<th>D21S11</th>
<th>D21S11</th>
<th>D7S820</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF01C</td>
<td>13</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AF02C</td>
<td>11</td>
<td>13</td>
<td>29</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td>AF03C</td>
<td>10</td>
<td>13</td>
<td>20</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

Uncovering Previous Mistakes During Data Review with Expert System

- 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-i3 “discordant” calls)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Types of Error Considered

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

FSS-i³ Rule Sets and Thresholds

Batch Summary

83 samples x 16 loci = 1,328 potential allele calls

Profile Results

- FP (full profile) = 1,136 times
- PP (partial profile) = 96 times
- FP-MIX (full profile with potential mixture) = 80 times
- NSD (no signal detected) = 16 times → negative control

Rules Fired

- No rules fired = 957 times
- Pref Amp A8 = 89 times
- High signal = 75 times
- Pull-up = 66 times
- Extra allele, etc. = 57 times
- Extra peak, etc. = 26 times
- Noise = 21 times
- Signal/Noise, etc. = 19 times
- Degradation = 13 times
- Peak Morph = 4 times
- Bin Offset = 1 time

Rules Fired by Locus

<table>
<thead>
<tr>
<th>Loc</th>
<th>#Total</th>
<th>#0</th>
<th>#1</th>
<th>#2+</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPOX</td>
<td>83</td>
<td>75</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>83</td>
<td>72</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>D16S539</td>
<td>83</td>
<td>68</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>D19S817</td>
<td>83</td>
<td>67</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>AMEL</td>
<td>83</td>
<td>66</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>D21S11</td>
<td>83</td>
<td>66</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>D5S818</td>
<td>83</td>
<td>65</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>FGA</td>
<td>83</td>
<td>64</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>D2S1338</td>
<td>83</td>
<td>63</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>D18S51</td>
<td>83</td>
<td>63</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>D7S800</td>
<td>83</td>
<td>57</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>TH01</td>
<td>83</td>
<td>56</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>D3S1358</td>
<td>83</td>
<td>52</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>D19S433</td>
<td>83</td>
<td>52</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>D8S1179</td>
<td>83</td>
<td>40</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>VWA</td>
<td>83</td>
<td>32</td>
<td>33</td>
<td>16</td>
</tr>
</tbody>
</table>

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 Sample

Only two samples had no rules fired

Provides a form of quality checks to the data examined

Concordance Evaluation

- Identifier data collected on ABI 3100; 83 samples processed in a single 96-well plate with a single allelic ladder (84 samples total)
- Typed with GeneScan/GenoTyper (MCK)
- Same data processed through GeneMapperID/FSS-i³ (JMB)
- Results from 81 samples compared (removed pos. & neg.):
  - 67 samples matched with no data review
  - 14 pairs exhibited a mismatch with unedited FSS-i³ results
- Examination of mismatches to determine which rules fired and if user would be able to make correct calls following editing

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
STR_MatchSamples.xls Output Under "Best Match" Showing Several Discordant Calls

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Loss” of an Allele</td>
<td>“Gain” of an Allele</td>
<td>“Wrong Call”</td>
</tr>
<tr>
<td>FSS-i³ Call (unedited)</td>
<td>Genotyper Call</td>
<td></td>
</tr>
<tr>
<td>D2S1338 7,24</td>
<td>vWA 16,18</td>
<td>FGA 7,51.2</td>
</tr>
<tr>
<td>D2S1338 24,26</td>
<td>vWA 18,18</td>
<td>FGA 22,26</td>
</tr>
</tbody>
</table>

Example 1

Unedited FSS-i³ Data:
D2S1338 Allele 26 “Loss”

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-i³ Data:
Pull-up from TH01 to vWA

FSS-i³ Rules Fired for vWA Locus

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

Correct vWA Type = 18,18

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

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-i3.

FSS-i3 Rules Fired

Spike in FGA High-Molecular Weight Region

 STR_MatchSamples.xls Output Under “Best Match” Showing Several Discordant Calls

---

Example 3

Unedited FSS-i3 Data: Spike in FGA High-Molecular Weight Region

Example 1

Example 2

Example 3

Unedited FSS-i3 Data:
Spike in FGA High-Molecular Weight Region

Correct FGA type = 22,26

FSS-i3 Rules Fired

Spike in FGA High-Molecular Weight Region

No pull up or peak morph. rule fired!— probably due to GeneMapperID data extraction problem?
If No Rules Fired, Were There Any Mistakes?

- 957 genotypes made with "no rules fired" (1,328 possible types across 83 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

Developmental Validation Studies?

Requests by NIST to both FSS and Promega for copies of developmental validation studies have not been acted upon...

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

What FSS i3 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-i3 v4.1 will supposedly open up capabilities

Thoughts Regarding FSS-i3 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 See Improved

Suggestions for next update:

- Detailed User Manual
- Capability of processing more kits and custom assays
- Accounting for all data points??
- 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
  - Accommodate for this problem in i-STReam mixture module even if allele calls cannot be made

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-i3 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-JJ-R-029 between NIJ and the NIST Office of Law Enforcement Standards

Software Installation at NIST:
Bob McLaren
Kim Huston
Curtis Knox

Training at Promega:
Bob McLaren
Kim Huston
Curtis Knox

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