NIST Resources for Validation

Carolyn R. (Becky) Hill and David Duewer
U.S. National Institute of Standards and Technology
Validation Webinar
August 6, 2014

New NIST Software Tools
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

SWGDAM Validation Guidelines

• Many of these programs can be used to help meet the SWGDAM Validation Guidelines for DNA Analysis Methods

Approved December 2012

Scientific Working Group on DNA Analysis Methods
Validation Guidelines for DNA Analysis Methods

NIST DNA Analyst Webinar Series:
Validation Concepts and Resources- Part 1
New NIST Software Tools
Developed by Dave Duewer (NIST)

From NIST STRBase Website:
- STR_AleleFreq
  - 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

New Programs Developed for CVP (not yet available on STRBase)
- Precision (Ladder_Precision)
- Known, nonprobative, and challenging samples (Concord_Revised)
- Sensitivity and stochastic (SensitivityAnalysis)
- Mixture studies (MixtureAnalysis)

NIST Software Tools and Validation
- NIST software programs can be used to assist in data analysis for both developmental and internal validation
- 3. Developmental Validation
  - 3.3 Sensitivity studies: SensitivityAnalysis
  - 3.5 Precision and accuracy: Ladder_Precision
  - 3.6 Case-type samples: Concord_Revised
  - 3.7 Population studies: STR_Genotype
  - 3.8 Mixture studies: MixtureAnalysis
  - 3.9 PCR-based studies: STR_AleleFreq (PHR) & STR_StutterFreq
NIST Software Tools and Validation

• 4. Internal Validation
  – 4.1 Known and nonprobative evidence samples or mock evidence samples: Concord_Revised
  – 4.2 Sensitivity and Stochastic Studies: SensitivityAnalysis
  – 4.3 Precision and accuracy: Ladder_Precision
  – 4.4 Mixture Studies: MixtureAnalysis
  – 4.5 Contamination assessment: Concord_Revised

Concordance Software: Concord_Revised

• This program processes known, nonprobative & case-type samples, & contamination assessment

Hit "DoAll" and it will prompt you to choose your Excel file and then process the data
*Must have comparative data

Concord_Revised

Calculates number of correct types & alleles, includes drop-in and drop-out

Provides summary of types by:
• Profile
• Worksheet
• Laboratory
• Locus
• Kit
Concord_Revised

Heat Map provided (created automatically by the program)

Summarizes drop-out and drop-in

Concord_Revised

Plots % Mis-Typings (drop-in and drop-out) vs. Average Allele Size (bp) to show trends

Sensitivity Software: SensitivityAnalysis

• This software processes data from sensitivity studies

Hit "DoAll" and it will prompt you to choose your Excel file and then process the data

*Program uses comparative data or it will output the consensus profile
Sensitivity Analysis

Summarizes % drop-in and drop-out for each locus at each concentration

Sensitivity Analysis

Plots Peak Height Ratio vs. Mean Peak Height (RFU) and PHR vs. Dilution Factor vs. Mean Peak Height (RFU) to show trends

Sensitivity Analysis

Provides Heat Maps for all sets of data (created automatically by the program)

Also provides Heat Map summaries (mean result) for each condition (i.e. kit & instrument)
Precision and Accuracy Software:
**Ladder_Precision**

- This software processes data from precision and accuracy studies

Hit "DoAll" and it will prompt you to choose your Excel file and then process the data

*Program uses allelic ladder data

---

Ladder_Precision

Program output for each condition (i.e. kit, instrument, data smoothing)

---

Ladder_Precision

- Allele Size (Standard Deviation) vs Allele size (bp)
- % Peak Height vs Allele Size (bp) or Peak Height (RFU)
- Peak Height (RFU) vs Allele Size (bp) and % Peak Area vs % Peak Height
Mixture Software: MixtureAnalysis

- This software processes data from mixture studies

Hit “DoAll” and it will prompt you to choose your Excel file and then process the data.

"Must have comparative data & focuses on drop-out of minor contributor"

MixtureAnalysis

Summarizes % drop-in and drop-out for each locus at each mixture ratio (similar to sensitivity program)

MixtureAnalysis

Provides Heat Maps for all sets of data (created automatically by the program)
Population Statistics Software:

**STR Genotype**

- This program processes population data/genotypes into allele frequencies and other population statistics for population studies.

Hit "Do It All" and it will prompt you to choose your Excel file and then process the data.

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

Population Data (N=568):

<table>
<thead>
<tr>
<th>Locus</th>
<th>Genotypes Observed</th>
<th>Alleles Observed</th>
<th>H(0bs)</th>
<th>PIC(exp)</th>
<th>PI</th>
<th>CumPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penta_E</td>
<td>113</td>
<td>20</td>
<td>0.8715</td>
<td>0.9023</td>
<td>0.0167</td>
<td>1.67E-02</td>
</tr>
<tr>
<td>D2S1338</td>
<td>66</td>
<td>13</td>
<td>0.8838</td>
<td>0.8850</td>
<td>0.0212</td>
<td>3.55E-04</td>
</tr>
<tr>
<td>D1S1656</td>
<td>85</td>
<td>15</td>
<td>0.8820</td>
<td>0.8840</td>
<td>0.0226</td>
<td>8.03E-06</td>
</tr>
<tr>
<td>D12S391</td>
<td>98</td>
<td>24</td>
<td>0.8785</td>
<td>0.8775</td>
<td>0.0255</td>
<td>2.16E-07</td>
</tr>
<tr>
<td>D18S51</td>
<td>78</td>
<td>19</td>
<td>0.8732</td>
<td>0.8731</td>
<td>0.0256</td>
<td>2.18E-07</td>
</tr>
<tr>
<td>Penta_D</td>
<td>70</td>
<td>16</td>
<td>0.8768</td>
<td>0.8535</td>
<td>0.0305</td>
<td>2.32E-04</td>
</tr>
<tr>
<td>D6S1043</td>
<td>91</td>
<td>24</td>
<td>0.8627</td>
<td>0.8475</td>
<td>0.0340</td>
<td>4.68E-05</td>
</tr>
<tr>
<td>D21S11</td>
<td>73</td>
<td>23</td>
<td>0.8468</td>
<td>0.8345</td>
<td>0.0372</td>
<td>1.60E-05</td>
</tr>
<tr>
<td>D19S433</td>
<td>70</td>
<td>16</td>
<td>0.8946</td>
<td>0.7914</td>
<td>0.0559</td>
<td>3.66E-15</td>
</tr>
<tr>
<td>D8S1179</td>
<td>70</td>
<td>20</td>
<td>0.8046</td>
<td>0.7914</td>
<td>0.0559</td>
<td>3.66E-15</td>
</tr>
<tr>
<td>vWA</td>
<td>33</td>
<td>10</td>
<td>0.8257</td>
<td>0.7863</td>
<td>0.0612</td>
<td>4.16E-18</td>
</tr>
<tr>
<td>D7S820</td>
<td>29</td>
<td>9</td>
<td>0.8063</td>
<td>0.7863</td>
<td>0.0612</td>
<td>4.16E-18</td>
</tr>
<tr>
<td>D16S539</td>
<td>26</td>
<td>7</td>
<td>0.7729</td>
<td>0.7396</td>
<td>0.0710</td>
<td>8.16E-18</td>
</tr>
<tr>
<td>AMEL</td>
<td>3</td>
<td>2</td>
<td>0.9965</td>
<td>0.3750</td>
<td>0.9930</td>
<td>7.92E-27</td>
</tr>
</tbody>
</table>

Observed Heterozygosity = H(0bs)
Expected Polymorphism Information Content = PIC(exp)
Probability of Identity (& cumulative) = PI and CumPI
Ms. Becky Hill

NIST DNA Analyst Webinar Series:
Validation Concepts and Resources - Part 1

D1S1656 Allele Frequencies

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>%</th>
<th>Thy</th>
<th>Adh</th>
<th>Allele %</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1S1656</td>
<td>11</td>
<td>71.5</td>
<td>16.7</td>
<td>8.8</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>12.9</td>
<td>6.5</td>
<td>12.3</td>
<td>18.4</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>11.7</td>
<td>8.6</td>
<td>16.6</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>11</td>
<td>16.6</td>
<td>16.6</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>14.3</td>
<td>7.5</td>
<td>6.5</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>10</td>
<td>16.7</td>
<td>16.7</td>
<td>12.6</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>15.3</td>
<td>6.9</td>
<td>3.1</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>16.7</td>
<td>11.2</td>
<td>6.9</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>14.3</td>
<td>11.1</td>
<td>4.6</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>16.7</td>
<td>11.2</td>
<td>6.9</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Calculates allele frequencies for all loci within the kit

Peak Height Ratio (PHR) Software:

**STR_AlleleFreq**

- This program processes peak height ratios (PHR) for PCR-based studies

Hit "Do It All" and it will prompt you to choose your Excel file and then process the data.

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

Balance Across All Loci
Interlocus Peak Balance

Peak Height Ratios

Heterozygote Peak Balance
Peak Height Ratios by Allele Spread

Heterozygote Peak Balance by Spread Between Alleles

<table>
<thead>
<tr>
<th>Locus</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Expected</th>
<th>2%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>D21S11</td>
<td>0.900</td>
<td>0.900</td>
<td>0.005</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td></td>
</tr>
<tr>
<td>D22S1043</td>
<td>0.920</td>
<td>0.920</td>
<td>0.007</td>
<td>0.915</td>
<td>0.920</td>
<td>0.920</td>
<td>0.920</td>
<td>0.920</td>
<td>0.920</td>
<td>0.920</td>
<td>0.920</td>
<td>0.920</td>
<td>0.920</td>
<td></td>
</tr>
</tbody>
</table>

Calculated separately for all loci

Stutter Software:
STR_StutterFreq

• This program processes stutter frequencies for PCR-based studies

Hit "Do It All" and it will prompt you to choose your Excel file and then process the data
http://www.cstl.nist.gov/strbase/software.htm
**Stutter Observed (Big Picture View)**

*D22S1045 stutter is in blue*

**Locus-Specific Stutter Values**

*Verified Max stutter amounts*

**D22S1045 Allele-Specific Stutter**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>Size</th>
<th>Stutter</th>
</tr>
</thead>
<tbody>
<tr>
<td>D22S1045</td>
<td>100</td>
<td>143</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>101</td>
<td>100</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>102</td>
<td>134</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>103</td>
<td>110</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>104</td>
<td>144</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>112</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Calculated separately for all loci
Stutter Trends (Size Scale)

*D22S1045 is in red

Stutter Trends (Repeat Scale)

*D22S1045 is in red

Stutter Density Plot

Median stutter = 6.57%
Median D22 stutter (in red) = 8.26%
Summary

• NIST has developed numerous software tools to assist in both developmental and internal validation studies
• All of these programs will be made available on STRBase for download when they are finalized
  – Peak Height Ratio, Stutter and Population Statistic programs are already available with example data sets
  http://www.cstl.nist.gov/strbase/software.htm

Acknowledgments

NIST Funding: Interagency Agreement 2008-DN-R-121 between the National Institute of Justice and NIST Office of Law Enforcement Standards

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 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.

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

NIST Team for This Work

Contact Info: becky.hill@nist.gov, 301-975-4275