NIST Resources for Validation

Carolyn R. (Becky) Hill and David Duewer
U.S. National Institute of Standards and Technology

Validation Webinar
August 6, 2014

NIST

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

New NIST Software Tools

Developed by Dave Duewer (NIST)

From NIST STRBase Website:

- **STR_AlleleFreq**
  - 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

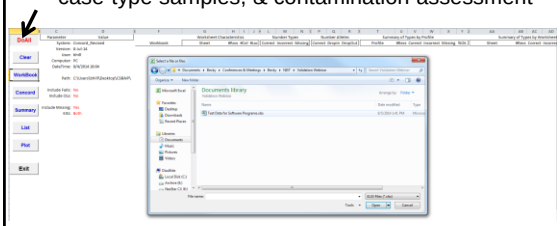
- NIST software programs can be used to assist in data analysis for both **developmental** and **internal** validation
- 3. Developmental Validation
  - 3.3 Sensitivity studies: SensitivityAnalysis
  - 3.5 Precision and accuracy: Ladder_Precision
  - 3.6 Case-type samples: Concord_Revised
  - 3.7 Population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

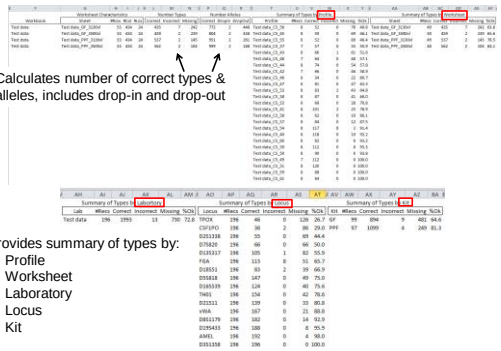
- This program processes known, nonprobative & case-type samples, & contamination assessment



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

*Must have comparative data

Concord_Revised



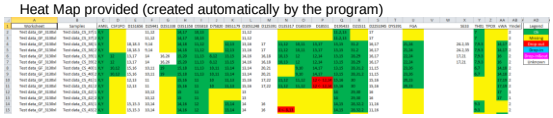
Calculates number of correct types & alleles, includes drop-in and drop-out

Provides summary of types by:

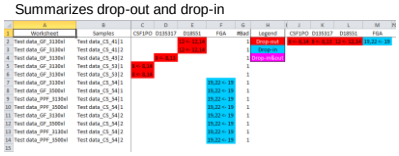
- Profile
- Worksheet
- Laboratory
- Locus
- Kit

Concord_Revised

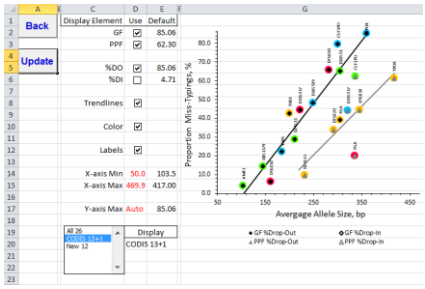
Heat Map provided (created automatically by the program)



Summarizes drop-out and drop-in



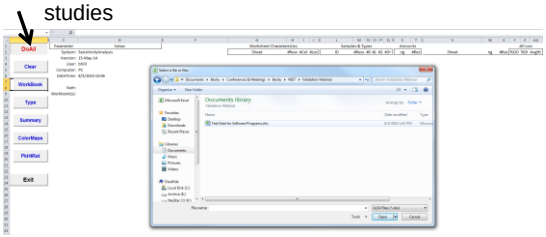
Concord_Revised



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

SensitivityAnalysis

	Worksheet Characteristics				Samples & Types					Amounts	
	Sheet	#Row	#Col	#Loc	ID	#Recs	#0	#1	#2	#3+	ng	#Rec
3	Test data_sensitivity_DLD.xlsx	Test data_Sens_PPF_3130xl	65	435	24	Test data_F1	64	2	8	16	4	24
4	Test data_sensitivity_DLD.xlsx	Test data_Sens_PPF_3500xl	65	435	24	Test data_M1	64		4	22	2	24
5	Test data_sensitivity_DLD.xlsx	Test data_Sens_GF_3130xl	65	435	24	Test data_M2	64		4	22	1	24
6	Test data_sensitivity_DLD.xlsx	Test data_Sens_GF_3500xl	65	435	24						0.5	24
7											0.25	24
8											0.125	24
9											0.0625	24
10											0.03125	24

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

SensitivityAnalysis

Back

Next

Display Element: Value

Worksheet Indx 1

Points ☑

BoxPlot ☑

Title ☑

PkHt Axis Max 27000

PkHt Max 8160

Worksheets

Indx	Name
1	Test data_Sens_GF_3130xl
2	Test data_Sens_GF_3500xl
3	Test data_Sens_PPF_3130xl
4	Test data_Sens_PPF_3500xl

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

SensitivityAnalysis

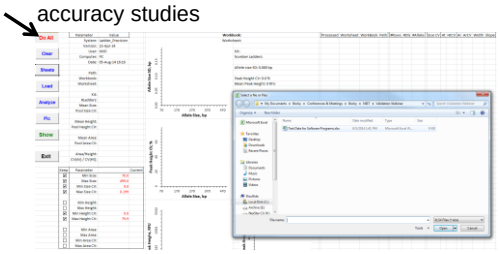
Provides Heat Maps for all sets of data (created automatically by the program)

Also provides Heat Map summaries (mean result) for each condition (i.e. kit & instrument)

Clicking on the box gives you the correct type

Precision and Accuracy Software: Ladder_Precision

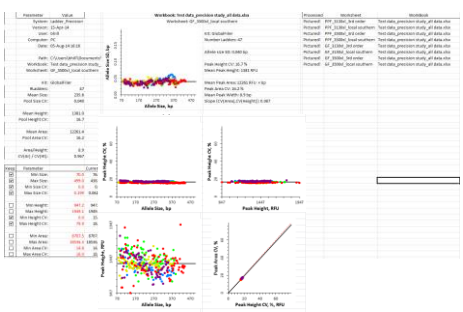
- This software processes data from precision and accuracy studies



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

*Program uses allelic ladder data

Ladder_Precision



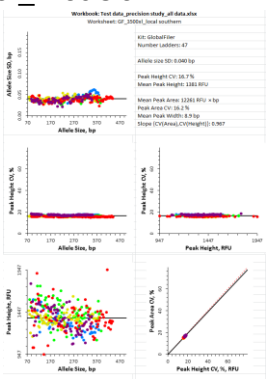
Program output for each condition (i.e. kit, instrument, data smoothing)

Ladder_Precision

Allele Size (Standard Deviation) vs Allele size (bp)

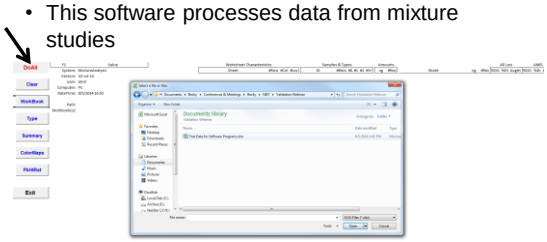
% Peak Height vs Allele Size (bp) or Peak Height (RFU)

Peak Height (RFU) vs Allele Size (bp) and % Peak Area vs % Peak Height



Mixture Software: MixtureAnalysis

- This software processes data from mixture studies

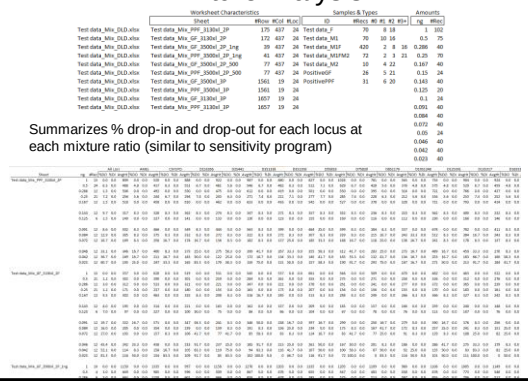


Hit "DoAll" and it will prompt you to choose your Excel file and then process the data

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

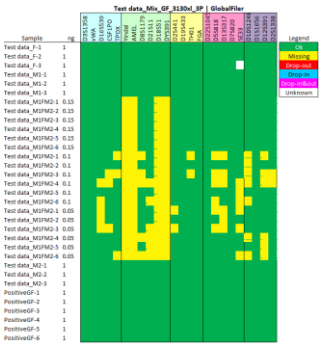
MixtureAnalysis

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

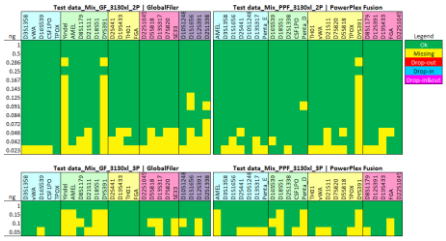


MixtureAnalysis

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



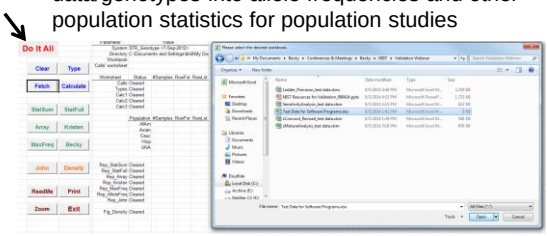
MixtureAnalysis



Also provides Heat Map summaries (mean result) for each condition (i.e. kit & instrument)

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)

Locus	Genotypes Observed	Alleles Observed	H(obs)	PIC(exp)	PI	CumPI
Penta_E	113	20	0.8715	0.9023	0.0167	1.67E-02
D2S1338	66	13	0.8838	0.8850	0.0212	3.55E-04
D1S1656	85	15	0.8820	0.8840	0.0226	8.03E-06
D12S391	98	24	0.8785	0.8755	0.0251	2.01E-07
D18S51	78	19	0.8732	0.8731	0.0258	5.19E-09
FGA	76	23	0.8873	0.8627	0.0299	1.55E-10
Penta_D	70	16	0.8768	0.8535	0.0335	5.20E-12
D6S1043	91	25	0.8627	0.8475	0.0343	1.78E-13
D21S11	73	23	0.8468	0.8345	0.0390	6.95E-15
D19S433	70	16	0.8046	0.7914	0.0569	3.96E-16
D8S1179	43	11	0.7764	0.7928	0.0580	2.30E-17
vWA	33	10	0.8257	0.7883	0.0612	1.41E-18
D7S820	29	11	0.8063	0.7686	0.0731	1.03E-19
D16S539	26	7	0.7729	0.7586	0.0760	7.81E-21
TH01	22	8	0.7500	0.7481	0.0804	6.28E-22
D13S317	29	8	0.7412	0.7415	0.0865	5.43E-23
D3S1358	24	10	0.7641	0.7321	0.0910	4.93E-24
CSF1PO	29	9	0.7394	0.7086	0.1021	5.04E-25
D5S818	32	9	0.7236	0.6959	0.1113	5.60E-26
TPOX	28	9	0.7201	0.6786	0.1243	6.97E-27
AMEL	3	2	0.9965	0.3750	0.9930	6.92E-27

Observed Heterozygosity = H(obs)
Expected Polymorphism Information Content = PIC(exp)
Probability of Identity (& cummulative) = PI and CumPI

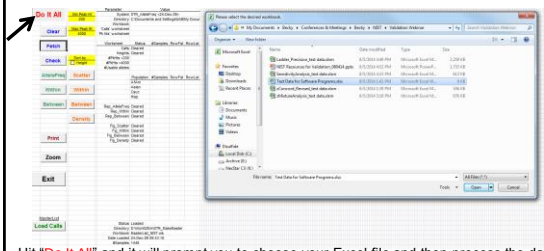
D1S1656 Allele Frequencies

Locus	Allele	Total #	Total %	Populations, % AfAm	Asian	Cauc	Hisp
D1S1656	10	12	0.9	1.6		0.4	0.7
(#allele: 15)	11	71	5.4	3.7	16.7	8.0	3.2
	12	126	9.6	6.5		12.3	10.4
	13	117	8.9	10.6		5.7	11.9
	14	192	14.6	23.7		7.1	12.2
	14.3	7	0.5	0.8		0.2	0.7
	15	208	15.8	17.6	50.0	14.9	12.9
	15.3	65	4.9	3.1		7.1	4.3
	16	167	12.7	11.0	16.7	12.6	15.8
	16.3	107	8.1	11.2		6.9	5.0
	17	47	3.6	2.5		4.6	3.6
	17.3	133	10.1	4.3		13.4	14.7
	18	5	0.4	0.2		0.6	0.4
	18.3	47	3.6	2.4	16.7	4.8	3.2
	19.3	14	1.1	0.8		1.3	0.7

Calculates allele frequencies for all loci within the kit

Peak Height Ratio (PHR) Software: STR_AlleleFreq

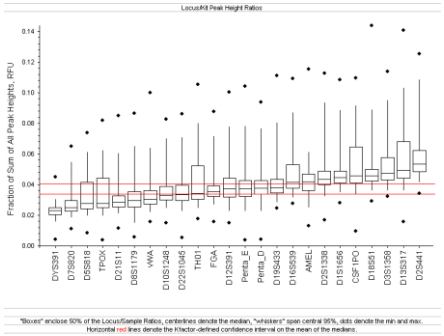
- This program processes peak height ratios (PHR) for PCR-based studies



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

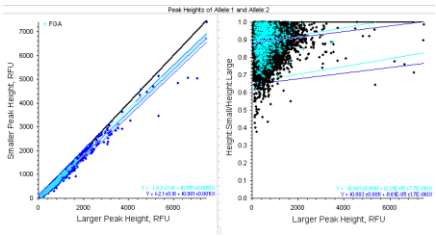


Interlocus Peak Balance

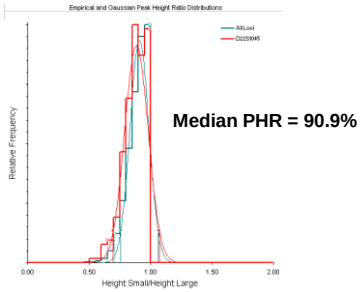
Locus/Kit Peak Height Ratios

		Mean		Median		Percentiles							
Locus	#	X	s(X)	X	s(X)	Min	2.5%	10%	25%	75%	90%	97.5%	Max
AMEL	662	0.042	0.010	0.042	0.008	0.013	0.025	0.030	0.036	0.047	0.052	0.061	0.116
CSF1PO	662	0.053	0.018	0.046	0.008	0.009	0.034	0.038	0.041	0.064	0.084	0.092	0.110
D10S1248	662	0.038	0.013	0.033	0.006	0.015	0.025	0.027	0.030	0.039	0.061	0.070	0.087
D12S391	657	0.041	0.014	0.037	0.009	0.015	0.023	0.027	0.032	0.044	0.061	0.078	0.100
D13S317	662	0.058	0.020	0.049	0.009	0.016	0.036	0.041	0.044	0.071	0.090	0.103	0.147
D16S539	661	0.049	0.017	0.042	0.008	0.029	0.032	0.034	0.038	0.053	0.078	0.087	0.109
D18S51	662	0.050	0.014	0.046	0.008	0.029	0.036	0.039	0.042	0.050	0.077	0.089	0.144
D19S433	661	0.044	0.015	0.038	0.006	0.025	0.029	0.032	0.035	0.043	0.071	0.081	0.111
D1S1656	662	0.048	0.014	0.045	0.005	0.028	0.035	0.038	0.041	0.049	0.069	0.089	0.108
D2S1	662	0.032	0.011	0.028	0.005	0.012	0.021	0.023	0.025	0.033	0.051	0.060	0.085
D22S1045	661	0.038	0.013	0.033	0.007	0.005	0.023	0.027	0.030	0.040	0.060	0.071	0.086
D2S1338	662	0.048	0.015	0.044	0.006	0.017	0.032	0.037	0.040	0.049	0.074	0.093	0.113
D2S441	661	0.062	0.020	0.054	0.009	0.034	0.042	0.045	0.048	0.062	0.099	0.108	0.125
D3S1358	662	0.055	0.018	0.047	0.008	0.032	0.036	0.040	0.043	0.058	0.088	0.095	0.114
D5S818	660	0.033	0.012	0.029	0.006	0.008	0.020	0.023	0.024	0.042	0.052	0.061	0.074
D7S820	660	0.029	0.010	0.025	0.005	0.011	0.019	0.021	0.023	0.030	0.047	0.055	0.070
D8S1179	661	0.033	0.013	0.030	0.007	0.006	0.015	0.020	0.025	0.035	0.053	0.065	0.087
DYS391	648	0.023	0.004	0.023	0.003	0.004	0.016	0.018	0.020	0.025	0.028	0.030	0.045
FGA	661	0.039	0.011	0.036	0.005	0.016	0.027	0.030	0.032	0.039	0.058	0.072	0.088
Penta_D	662	0.041	0.013	0.038	0.007	0.004	0.023	0.030	0.034	0.043	0.062	0.077	0.107
Penta_E	662	0.040	0.014	0.037	0.008	0.004	0.022	0.028	0.032	0.043	0.055	0.078	0.104
TH01	661	0.041	0.016	0.034	0.008	0.018	0.024	0.027	0.030	0.052	0.067	0.080	0.105
TPOX	659	0.034	0.013	0.028	0.006	0.004	0.020	0.022	0.025	0.044	0.054	0.062	0.082
vWA	661	0.034	0.012	0.030	0.006	0.016	0.022	0.025	0.027	0.036	0.054	0.064	0.100

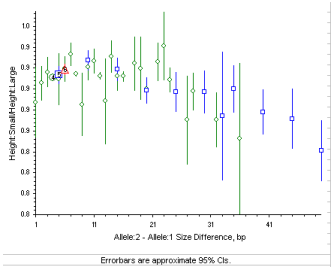
Peak Height Ratios



Heterozygote Peak Balance



Peak Height Ratios by Allele Spread



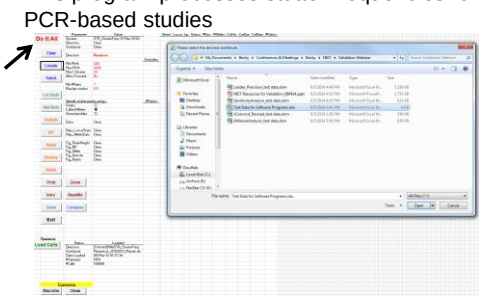
Heterozygote Peak Balance by Spread Between Alleles

			Mean		Median		Percentiles							
Locus	Δbp	#	X	s(X)	X	s(X)	Min	2.5%	10%	25%	75%	90%	97.5%	Max
D12S391	4	149	0.896	0.076	0.911	0.077	0.684	0.730	0.795	0.841	0.954	0.982	0.992	1.000
	7	10	0.923	0.074	0.954	0.078	0.780	0.804	0.875	0.905	0.957	0.967	0.968	0.968
	8	111	0.895	0.074	0.913	0.080	0.586	0.727	0.803	0.838	0.964	0.978	0.991	0.997
	12	93	0.901	0.073	0.916	0.074	0.653	0.745	0.809	0.856	0.957	0.986	0.995	0.998
	16	58	0.918	0.073	0.930	0.073	0.765	0.788	0.836	0.873	0.972	0.987	0.992	0.997
	20	57	0.893	0.077	0.902	0.081	0.715	0.734	0.788	0.846	0.959	0.984	0.995	0.999
	24	24	0.879	0.081	0.903	0.068	0.648	0.668	0.763	0.838	0.943	0.971	0.988	0.999
	28	15	0.918	0.069	0.930	0.058	0.765	0.779	0.816	0.907	0.957	0.982	0.994	0.996
	32	11	0.896	0.072	0.889	0.089	0.791	0.794	0.804	0.829	0.958	1.000	1.000	1.000

Calculated separately for all loci

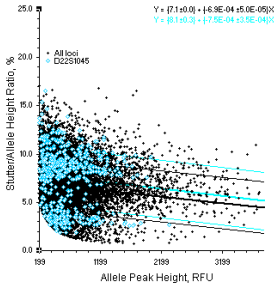
Stutter Software: STR_StutterFreq

- 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

Created on: 23-May-12 16:37:36
From Version: 29-Feb-12

Filter	Setting
Fetch:	Stored profiles
Min PHR:	200
Max PHR:	4000
Max %Stutter:	30
Allow Forwards:	No
Min #Ratio:	7
Bp/vln:	0.50

Locus	#Ratio	Mean	SD	Median	MADe	Percentage Stutter Min	1%	2.5%	5%	10%	25%	75%	90%	95%	97.5%	99%	Max
THO1	432	2.5	1.0	2.2	1.0	0.9	1.1	1.2	1.3	1.4	1.7	3.2	4.0	4.5	5.2	5.8	10.0
Penta_D	427	2.9	1.6	2.5	1.2	0.7	0.8	1.0	1.1	1.3	1.8	3.5	4.9	5.9	6.4	7.4	12.7
TPOX	452	3.3	1.3	3.2	1.5	0.9	1.5	1.6	1.7	1.8	2.2	4.1	4.9	5.8	6.3	6.9	9.6
Penta_E	685	4.0	2.4	3.5	1.6	0.7	0.9	1.1	1.3	1.7	2.6	4.7	6.8	8.8	15.3	12.8	15.3
D7S820	429	5.2	1.7	4.9	1.6	1.8	2.5	2.7	2.8	3.1	4.0	6.2	7.5	8.7	9.1	10.3	11.2
D2S441	78	5.1	1.9	5.1	1.5	1.9	1.9	2.3	2.7	3.2	4.0	6.1	6.9	7.7	8.7	9.7	14.7
D13S317	526	5.1	1.9	5.2	1.7	1.2	1.7	1.8	2.0	2.5	4.0	6.2	7.4	8.1	8.8	10.0	17.1
D5S818	388	5.8	1.6	5.9	1.3	1.8	2.6	3.0	3.4	3.8	4.9	6.7	7.6	8.2	9.1	9.8	10.7
CSF1PO	452	6.1	1.6	6.0	1.4	1.6	2.4	2.8	3.4	4.2	5.2	7.0	8.3	9.0	9.8	10.6	12.4
D16S539	496	6.5	1.7	6.3	1.7	2.3	3.6	3.8	4.0	4.4	5.3	7.6	8.9	9.5	10.3	10.8	14.3
D19S433	589	6.6	1.6	6.5	1.4	1.8	3.3	3.6	4.3	4.7	5.6	7.5	8.6	9.4	10.2	10.7	12.9
D6S1043	520	6.8	1.8	6.7	1.4	2.5	3.3	3.8	4.3	5.1	5.8	7.8	8.6	9.4	10.2	10.9	11.8
vWA	461	6.9	2.3	6.8	1.9	0.9	1.2	1.5	2.3	4.3	5.6	8.2	9.6	10.4	11.4	12.3	28.6
FGA	644	6.9	1.7	6.8	1.7	3.4	3.7	3.9	4.2	4.7	5.7	8.1	9.2	9.8	10.3	11.0	12.6
D18S51	453	8.0	1.8	7.7	1.6	3.7	4.7	5.0	5.5	6.0	6.8	8.9	10.5	11.4	12.0	13.2	15.3
D21S11	633	8.1	1.6	8.0	1.4	4.1	4.9	5.3	5.7	6.1	7.1	9.0	10.2	10.8	12.0	12.9	15.3
D3S1358	873	8.2	2.0	8.1	2.2	2.3	4.2	4.9	5.3	5.8	6.7	9.7	10.9	11.5	12.3	12.8	15.6
D22S1045	552	7.6	3.0	8.3	2.7	1.8	2.2	2.4	2.8	3.0	4.5	9.6	11.2	11.7	12.4	13.5	16.8
D12S391	696	8.6	2.7	8.3	3.0	3.7	4.0	4.3	4.6	5.0	6.5	10.5	12.4	13.2	14.3	15.3	16.8
D8S1179	745	8.6	2.2	8.4	2.3	3.6	4.3	4.8	5.1	5.4	6.8	9.8	11.4	12.4	13.1	13.6	16.6
D2S1338	834	8.7	1.9	8.5	2.0	4.0	5.4	5.6	5.9	6.3	7.2	9.9	11.2	11.8	12.7	13.4	14.9
D19S433	494	8.8	1.6	8.5	1.5	3.1	5.4	5.9	6.2	6.7	7.5	9.5	10.6	11.2	11.9	12.6	16.1

Settings Used for Analysis

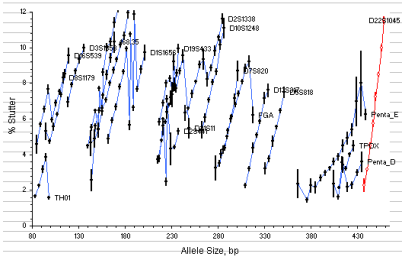
*Verified Max stutter amounts

D22S1045 Allele-Specific Stutter

Locus	Allele	Size	Stutter #	Median	MADe
D22S1045	10	433.9	12	2.3	0.6
	11	437.1	111	3.2	0.5
	12	440.2	34	4.5	0.5
	13	443.4	8	5.8	0.2
	14	446.5	44	7.3	0.7
	15	449.7	186	8.5	0.7
	16	452.7	108	10.0	0.9
	17	455.6	44	11.5	0.6
		Avg	547	6.6	0.6
		SD		3.3	

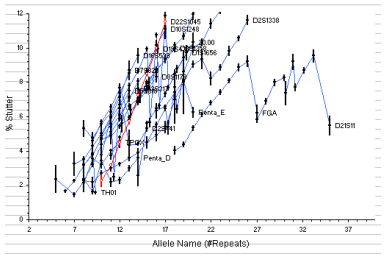
Calculated separately for all loci

Stutter Trends (Size Scale)



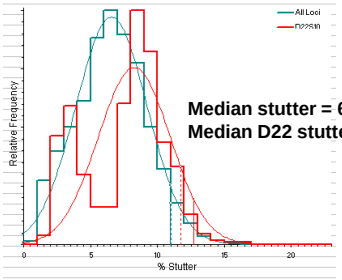
*D22S1045 is in red

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

Dave Duewer

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