http://www.cstl.nist.gov/div831/strbase
david.duewer@nist.gov

New Tools to Aid Work with STR Profile Mixtures:
mixSTR and Virtual MixtureMaker

David L. Duewer[1], Cecelia A. Crouse[2], and John M. Butler[1]

[1]National Institute of Standards and Technology (NIST), Gaithersburg, MD; [2]Palm Beach County Sheriff's Office (PBSO) Crime Laboratory, West Palm Beach, FL

A number of Microsoft Excel-based tools have been developed at NIST to aid internal research projects funded by the National Institute of Justice. Some of these software tools may prove useful for application in forensic casework or validation involving STR mixture interpretation. Two programs are being made available for download from the NIST STRBase website.

The ***mixSTR system***, developed in collaboration with the Palm Beach County Sheriff's Office, helps identify reference profiles in sets of mixed-source samples. Given allele lists for one or more evidentiary samples and all relevant suspect and control reference profiles, the mixSTR system provides detailed lists of the match and mismatch alleles among all samples and STR profiles. Sorted summaries of the percent of included alleles and loci for all comparisons are also provided. This tool is intended for use by forensic scientists working with mixed-source samples particularly cases involving a number of suspects.

Virtual MixtureMaker was developed to aid in sample selection for the recent NIST Mixture Interpretation Study (MIX05). This program performs a pairwise comparison of STR profiles in a dataset and reports the number of loci possessing 0,1,2,3,4,5, or 6 alleles in all possible two-person mixtures.

mixSTR Program

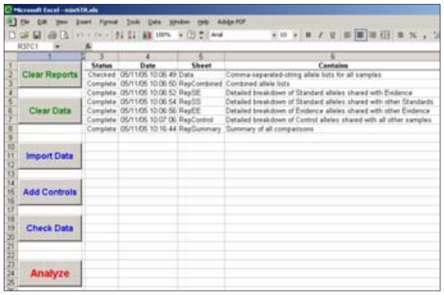
- Does not interpret data (relies on user inputted alleles following STR data review)
- Aids in the organization of STR mixture information
- Considers only the presence/absence of alleles (no peak heights used)

ControlPanel

The "ControlPanel" worksheet within this Excel-based program has a panel of buttons that allow you to drive the mixSTR system. The current status of the data and report sheets are also displayed.

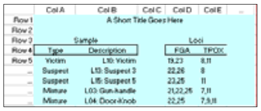
Currently there are six functions in the mixSTR ControlPanel:

Clear Reports: Clears all contents of all the report sheets.
Clear Data: Clears all contents of the Data sheet.
Import Data: Clears any old data and imports data from "Merged Landscape"-style Excel workbooks.
Add Controls: Adds Control samples from specified Excel sheets. Control alleles must be specified as comma-separated lists.
Check Data: Checks the Data sheet information compliance with mixSTR format.
Analyze: Clears any old reports, checks the data on the "Data" worksheet and, if all is ok, produces all the Reports.



Status	Date	Sheet	Contains
Checked	05/11/05 10:06:49	Data	Comma-separated-string allele lists for all samples
Complete	05/11/05 10:06:50	RepCombined	Combined allele lists
Complete	05/11/05 10:06:52	RepSE	Detailed breakdown of Standard alleles shared with Evidence
Complete	05/11/05 10:06:54	RepSS	Detailed breakdown of Standard alleles shared with other Standards
Complete	05/11/05 10:06:56	RepEE	Detailed breakdown of Evidence alleles shared with other Evidence
Complete	05/11/05 10:07:06	RepControl	Detailed breakdown of Control alleles shared with all other samples
Complete	05/11/05 10:10:44	RepSummary	Summary of all comparisons

Example of data format required for input



	Col A	Col B	Col C	Col D	Col E	…
Row 1	A Short Title Goes Here					
Row 2						
Row 3	Sample			Loci		
Row 4	Type	Description		FGA	TPOX	
Row 5	Victim	L10: Victim		19,23	8,11	
…	Suspect	L13: Suspect 3		22,26	8	
…	Suspect	L15: Suspect 5		23,25	11	
…	Mixture	L03: Gun-handle		21,22,25	7,11	
…	Mixture	L04: Door-Knob		22,25	7,9,11	

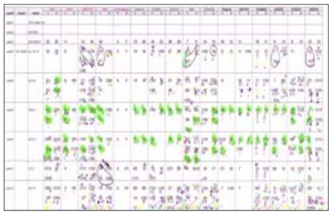
Comparisons are made between suspect and evidence (S/E) alleles, suspect and suspect (S/S) alleles (to look for potential close relatives), evidence and other evidence (E/E) sample(s) alleles (to see how various evidentiary samples compare to one another), and controls to evidence (C/E) and controls to suspect (C/S) alleles (as a quality control contamination check).

Use in a real case…

Courtesy of Catherine Cothran from PBSO

Working with multiple suspects and multiple mixtures manually can be mind-numbing…

Manual comparisons

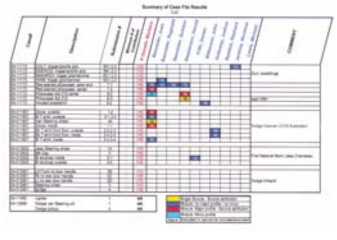


PBSO data using a color or symbols to correlate results from each locus with each piece of evidence to try and determine at how many loci alleles from each of the suspects were present

mixSTR S/E output

Example of suspect to evidence (S/E) comparisons made in this case. Note that the suspect is 21,23 at FGA while the evidence contains 23,24* (* indicates that allele 24 is a minor component). Thus this suspect has allele 23 in common and is missing allele 24 in the evidence.

Summary data

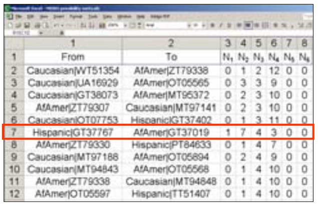


Presentation used in court to describe data from this case. The mixSTR program helped determine the information presented in this chart to show the presence of alleles from multiple suspects in multiple pieces of evidence.

Programs can be downloaded from NIST STRBase web site:
http://www.cstl.nist.gov/div831/strbase/software.htm

Virtual MixtureMaker

- Creates mixture combinations through pairwise comparisons of input STR profiles
- Returns information on the number of loci possessing 0,1,2,3,4,5, or 6 alleles in each 2-person mixture (also reports number of loci in each sample with 0,1,2, or 3 alleles)
- Useful for selection of samples in mixture or validation studies with various degrees of overlapping alleles in combined STR profiles
- Useful in checking for potentially related individuals in a population database



	1	2	3	4	5	6	7	8
1	From	To	N_1	N_2	N_3	N_4	N_5	N_6
2	Caucasian|WT51354	AfAmer|ZT79338	0	1	2	12	0	0
3	Caucasian|UA16929	AfAmer|OT05565	0	3	3	9	0	0
4	Caucasian|GT38073	AfAmer|MT95372	0	2	3	10	0	0
5	AfAmer|ZT79307	Caucasian|MT97141	0	2	3	10	0	0
6	Caucasian|OT07753	Hispanic|GT37402	0	1	3	11	0	0
7	Hispanic|GT37767	AfAmer|GT37019	1	7	4	3	0	0
8	AfAmer|ZT79330	Hispanic|PT84633	0	1	4	7	0	0
9	Caucasian|MT97188	AfAmer|OT05894	0	2	4	9	0	0
10	Caucasian|MT94843	AfAmer|OT05568	0	1	4	10	0	0
11	AfAmer|ZT79338	Caucasian|MT94848	0	1	4	10	0	0
12	AfAmer|OT05597	Hispanic|TT51407	0	1	4	10	0	0

Based on comparison of NIST 700 U.S. population samples examined with 15 STR loci (Identifiler data)

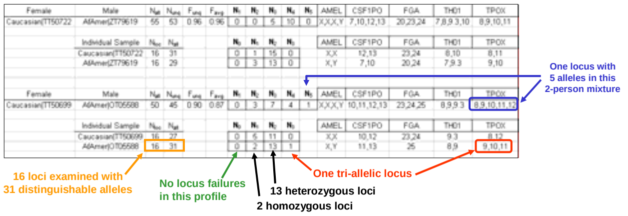
When the STR profiles for these two individuals are combined to create a 2-person mixture, the mixture profile will contain 1 locus with a single allele, 7 loci with two alleles, 4 loci with three alleles, and 3 loci with four alleles (and no loci with 5 or 6 alleles, which is only possible if one or both samples possess tri-allelic patterns at the same STR locus).

For sample selection in the recent NIST Mixture Interpretation Study (MIX05), we used *Virtual MixtureMaker* to:

1) Generate all possible two person mixtures through comparing 40 females against 660 males tested previously with Identifiler (15 STR loci); see Butler et al. (2003) *J. Forensic Sci.* 48(4):908-911; data available on STRBase at http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/JFS2003IDresults.xls

2) Sort all virtual mixtures based on number of loci possessing 1, 2, 3, 4, or 5 alleles with a two-person mixture

We then selected sample combinations exhibiting a variety of allele combinations and mixed the samples in the lab at predefined ratios to create the STR mixture profiles for the MIX05 study (see http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm).



Female	Male	N_{all}	N_{unq}	F_{unq}	F_{avg}	N_1	N_2	N_3	N_4	N_5	AMEL	CSF1PO	FGA	TH01	TPOX
Caucasian|TT50722	AfAmer|ZT79619	55	53	0.96	0.96	0	0	5	10	0	X,X,X,Y	7,10,12,13	20,23,24	7,8,9.3,10	8,9,10,11

Individual Sample	N_{loc}	N_{all}	N_0	N_1	N_2	N_3	AMEL	CSF1PO	FGA	TH01	TPOX
Caucasian|TT50722	16	31	0	1	15	0	X,X	12,13	23,24	8,10	8,11
AfAmer|ZT79619	16	29	0	3	13	0	X,Y	7,10	20,24	7,9.3	9,10

Female	Male	N_{all}	N_{unq}	F_{unq}	F_{avg}	N_1	N_2	N_3	N_4	N_5	AMEL	CSF1PO	FGA	TH01	TPOX
Caucasian|TT50699	AfAmer|OT05588	50	45	0.90	0.87	0	3	7	4	1	X,X,X,Y	10,11,12,13	23,24,25	8,9,9.3	8,9,10,11,12

Individual Sample	N_{loc}	N_{all}	N_0	N_1	N_2	N_3	AMEL	CSF1PO	FGA	TH01	TPOX
Caucasian|TT50699	16	27	0	5	11	0	X,X	10,12	23,24	9.3	8,12
AfAmer|OT05588	16	31	0	2	13	1	X,Y	11,13	25	8,9	9,10,11

One locus with 5 alleles in this 2-person mixture

16 loci examined with 31 distinguishable alleles

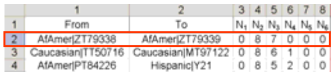
No locus failures in this profile

13 heterozygous loci

2 homozygous loci

One tri-allelic locus

Virtual MixtureMaker can also be used **for screening STR profiles in population databases for the presence of individuals who may be closely related and thus share many alleles** (i.e., have many 2 allele loci and few 4 allele loci when pairwise STR profiles are compared). Using this allele comparison approach with 15 autosomal STR loci (Identifiler data), we identified in our U.S. dataset of 700, two individuals (ZT79338 and ZT79339) who we believe to be brothers. Since our NIST population samples were provided anonymously from several blood banks, we have no way to go back and verify with the original sample sources. However, the single digit difference in the numbering of the two samples likely indicates that these samples were collected at the same site in sequential fashion and thus could be members of the same family. In addition to the high degree of allele sharing with autosomal STRs, we have found that these two samples match at 46 of 47 Y-STRs examined and have identical mtDNA sequences over the entire ~1100 bp control region.



	1	2	3	4	5	6	7	8
1	From	To	N_1	N_2	N_3	N_4	N_5	N_6
2	AfAmer|ZT79338	AfAmer|ZT79339	0	8	7	0	0	0
3	Caucasian|TT50716	Caucasian|MT97122	0	8	6	1	0	0
4	AfAmer|PT84226	Hispanic|Y21	0	8	5	2	0	0

Putative brothers *(46/47 Y-STRs examined match; mtDNA 1100 bp control region sequences are identical)*

NIST Interlaboratory Mixture Interpretation Study 2005 (MIX05)

This interlaboratory study was designed to evaluate (1) the current "lay of the land" regarding STR mixture interpretation across the forensic DNA typing community and (2) to aid development of training tools to enable improved mixture interpretation and more uniform reporting of mixtures in the future.

From January-March 2005 (with data still accepted until end of June 2005)**, over 60 different laboratories evaluated the same mixture data from four different case scenarios using electrophoretic data generated at NIST with multiple STR kits. As of June 2005, these data are being reviewed to understand the various approaches to mixture interpretation.**

In each case, NIST provided the "evidence" sample result (from a mock sexual assault— an "intimate sample"), which was a mixture of one male "perpetrator" and one female "victim". The "victim" reference sample was also provided for each scenario. Results were produced from the same lot of DNA mixtures using five different STR kits: Profiler Plus, COfiler, SGM Plus, Identifiler, and PowerPlex 16. ABI 3100 electropherograms from these kits are available for download from the NIST STRBase website (http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm) and may be used for training purposes with either GeneScan/Genotyper or GeneMapper*ID* data review.

Another benefit of the MIX05 interlab study is that interpretation guidelines from many laboratories are being compared to one another for the first time in an effort to determine challenges facing development of "expert systems" for automated mixture interpretation.

Acknowledgments and Disclaimer

Margaret Kline from NIST and Catherine Cothran from PBSO provided valuable input into these programs. These projects were funded by the National Institute of Justice through interagency agreement 2003-IJ-R-029 to the 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.