Evaluation of Inter-laboratory Probabilistic Genotyping Parameters

or: How I Learned to Stop Worrying and Love PG Software

Safia Boodoosingh, Hannah Kelly, James Curran, Tim Kalafut
tim.kalafut@shsu.edu

Department of Forensic Science
Sam Houston State University
Huntsville, TX, USA

Sam Houston State University

1

STRmix™ and variance values

- Early versions used static variance values
  - DyNAmix and v1.06 were really early versions
  - Online with v2.06 in November of 2014
  - We saw some struggles with low quality data
- Starting in v2.3, STRmix™ "varies the variance"
  - Low quality data typically uses a "forgiving" variance
    - Allows for poorer peak height ratios between heterozygote alleles
  - High quality data can have a "picky" variance
    - Requires better phr for hets

2

2

Variance is part of validation

- Model Maker mode of STRmix™ models the variance typically present in data
- Based on a gamma distribution
  - Larger mode values, the looser the variance
  - Smaller mode values, the tighter the variance
- Variance used for deconvoluting a casework sample varies
- Good phr has smaller variance, lower phr has larger
  - (A bit simplistic, but it works)

3

3

How much "variance" can STRmix™ handle?

- Variance is determined in Model Maker…
- …but is then optimized for each sample…
- …can STRmix™ run just fine for Lab A using Lab B's settings?

- We reached out to labs looking for participants
  - 8 labs agreed to participate
  - Each gave us twenty 2-, 3-, and 4-person **mixture input files**
  - A range of templates and ratios
  - Provided their STRmix™ kits/stutter files

4

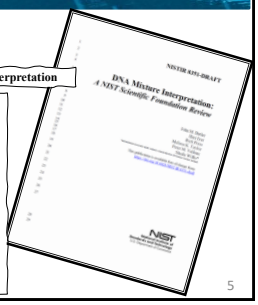
4

In the meantime…

- The NIST draft report on the status of mixtures came out

Chapter 4: Reliability of DNA Mixture Measurements and Interpretation

In this report, we divide the challenges presented by DNA mixtures into two main categories. The first involves the *reliability* of mixture interpretation methods when used with DNA evidence of varying complexity. […] In this report, we use the "plain English" definition of reliability as a measure of trustworthiness. A highly reliable method is one that consistently produces accurate results. Reliability is not a yes or no question, but a matter of degree. Understanding the degree of reliability of a method can help the user of that information decide whether they should trust the results of that method when making important decisions.



5

5

Key Takeaways

KEY TAKEAWAY #4.1: Reliability(?) – Validation studies, interlaboratory studies

KEY TAKEAWAY #4.4: Comparable(?) reliability – Need established criteria, and acceptable level of reliability

KEY TAKEAWAY #4.6: Variability and uncertainty – Different analysts and different laboratories

KEY TAKEAWAY #4.7: Degree of reliability – validation studies, known samples similar to casework

6

6

Significance of Research

- Two goals at the start:
  - Look at how robust/consistent/reliable STRmix™ is as a tool
  - Look at the possibility of labs doing direct sharing of data and analyzing one another's case samples
- Scope of project
  - 115 mixtures; 8 labs; 6,120 STRmix™ decons, >61 million LRs
  - Manuscript is in preparation
- Bonus:
  - Direct response to NIST Scientific Foundation Review report for DNA Mixture Interpretation

7

7

Research Questions

- Robustness – "Reliable" per NIST?
- Does STRmix™ produce similar LRs for ground truth donors using "non-validated" parameters?
  - NIST didn't define reliability metric or threshold
  - "Very close LRs" might help both define and demonstrate reliability
- Alternate way to share data – "Accurate" per NIST?
- Can ground truth donors from a mixture be correctly identified using "non-cognate" parameters?
  - NIST did not define accuracy metric or threshold
  - Similar ranking compared to LRs of non-donors might help

8

8

Materials and methods – "Robustness"

LR using **Lab A** STRmix™ settings on **Lab A** data

COMPARISON

LR using **Lab B** STRmix™ settings on **Lab A** data

Similar

Different

9

9

Materials and methods – "Accuracy"

DAVE (from Lab A)

LAB A KIT

Dave has the highest LR (or Xth on the list)

LAB B KIT

Does Dave still have the highest LR?? (or Xth on the list)

10

10

Database searching to find Dave

- We created a database of 10,000 non-donors
- We seeded it with the 82 ground truth donors
  - Actually, only 81 donors – one known donor with two different designations was found
- We ran "decons" followed by "Database Search" to get the LRs

11

11

STRmix™ Validation

- Stutter validation
  - Use single source samples to determine expected stutter ratios
  - Done outside of STRmix™
    - "Normal" back and forward stutter
    - "Exotic" stutter – double back stutter (-8 bp), half stutter (-2 bp)
    - Possibly -6 bp or +2 bp
- Use Model Maker to characterize "variance"
  - Peak variance between expected and observed peak heights
    - Proxy for peak height ratio
  - Stutter variance between expected and observed stutter peaks

12

12

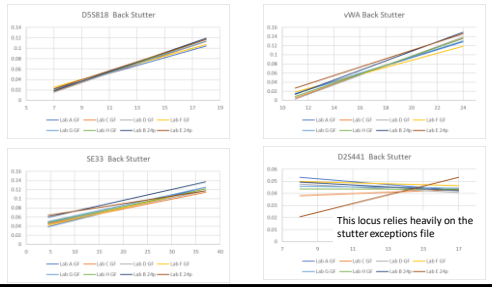
Two stutter files needed

- Develop stutter data files that live inside STRmix™
- Stutter exceptions
  - Just a data table of stutter values for alleles at loci
  - Boring to look at
- Slope of a line
  - For loci with a simple linear STR; or for unseen alleles at a locus
  - Can be pretty to look at

13

13

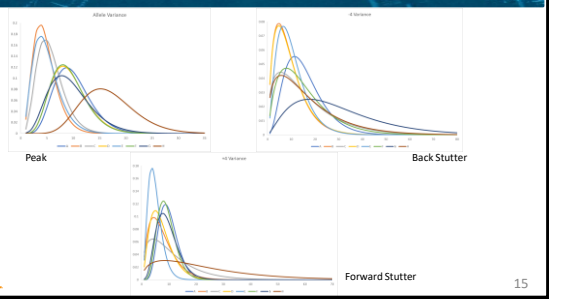
Stutter comparison across labs



14

14

Variance differences between the 8 labs



15

15

Additional validation issues/differences

- Analytical thresholds (AT)
  - Kit specific?
  - Dye channel specific?
  - Locus specific?
- Which kit to use in the first place
- How many PCR cycles?

16

16

Analytical Thresholds

- Validated by each lab (All labs used 3500 instruments)

Lab	Analytical threshold for each channel (RFU)				
	Blue	Green	Yellow	Red	Purple
A	50	65	45	55	60
B	45	55	65	75	45
C	40	40	40	40	40
D	80	60	70	60	60
E	60	60	60	60	60
F	75	75	75	75	75
G	200	200	200	200	200
H	100	100	100	100	100

Labs A, B, and D used locus specific thresholds

Labs C, E, F, G and H used a single threshold

17

17

But let's up the level of difficulty…

- We had both GlobalFiler™ and Investigator® 24plex labs
  - They have the same autosomal loci
  - Different order for loci and base pair sizes for alleles
- We had labs that used both 28 and 29 PCR cycles
  - Both kits had both cycles across the study

18

18

Kits, cycles, and native STRmix™ version

Lab Name	PCR Kit	No. of PCR Cycles	STRmix™ version
A	GlobalFiler™	28	2.5.11
B	Investigator® 24Plex QS	28	2.6.3
C	GlobalFiler™	28	2.6.3
D	GlobalFiler™	29	2.6.0
E	Investigator® 24Plex QS	29	2.4.06
F	GlobalFiler™	29	2.5.11
G	GlobalFiler™	29	2.4.06
H	GlobalFiler™	29	2.7.0
SHSU	Adapted all kits for v2.9(.1)		2.9(.1)

19

19

Deconvolution conditioning strategy

NOC	Propositions H1	Propositions H2	POI
2p	1+2	U+U	1, 2
	1+2	1+U	2
	2+1	2+U	1
3p	1+2+3	U+U+U	1, 2, 3
	1+2+U	1+U+U	2
	1+3+U		3
	2+1+U	2+U+U	1
	2+3+U		3
	3+1+U	3+U+U	1
	3+2+U		2
	1+2+3	1+2+U	3
	1+3+2	1+3+U	2
	2+3+1	2+3+U	1
4p	1+2+3+4	U+U+U+U	1, 2, 3, 4
	1+2+3+4	1+2+3+U	4
	1+2+4+3	1+2+4+U	3
	1+3+4+2	1+3+4+U	2
	2+3+4+1	2+3+4+U	1

2-person mixtures used 3 decons

3-person mixtures used 7 decons

4-person mixtures used 5 decons

20

20

Conditioning

A18: 100 pg amp;
4-person 10:5:2:1

6.46 x 10^{-1} → CONDITIONING → 1.14 x 10^{4}

$$LR = \frac{D + U + U + U}{U + U + U + U}$$

True Donor LR rank is #68
45 Random Donors LR > 1

$$LR = \frac{D + K1 + K2 + K3}{U + K1 + K2 + K3}$$

True Donor LR rank is #1
28 Random Donors LR > 1

SPOILER ALERT: *CONDITIONING* has a much bigger effect on the LR than anything else we tested

21

21

Comparison of extreme AT (40 vs 200)

Lab C analyzed at 40 rfu, Lab G analyzed at 200 rfu

Mixture No.	STRmix™ Kit	Donor 1 *LR*	Donor 1 Rank
C20 4p 1-1-1-1 0.2ng 40rfu AT	A	1.79E+06	4
	B	1.76E+06	4
	C	2.09E+06	3
	D	1.61E+07	2
	E	8.05E+04	4
	F	1.31E+06	4
	G	1.41E+06	4
	H	6.84E+06	2

22

22

Comparison of 28 to 29 cycles

Lab A used 28 cycles, Lab D used 29 cycles (both GlobalFiler)

Mixture No.	STRmix™ Kit	Major Donor *LR*	Major Donor Rank
A11 3p 10:5:1 0.1 ng	A	1.92E+21	1
	B	1.90E+22	2
	C	5.74E+21	2
	D	1.03E+21	1
	E	1.89E+22	2
	F	1.22E+21	1
	G	2.61E+20	1
	H	1.82E+18	1

23

23

Comparison of GlobalFiler to Investigator 24plex

Lab E used Investigator 24plex, Lab F used GlobalFiler (both 29 cycles)

Mixture No.	STRmix™ Kit	Major Donor *LR*	Major Donor Rank
E12 3p 10:1:1 1.0 ng	A	1.50E+30	1
	B	1.60E+30	1
	C	1.81E+30	1
	D	1.88E+30	1
	E	1.63E+30	1
	F	1.65E+30	1
	G	1.56E+30	1
	H	1.90E+30	1

24

24

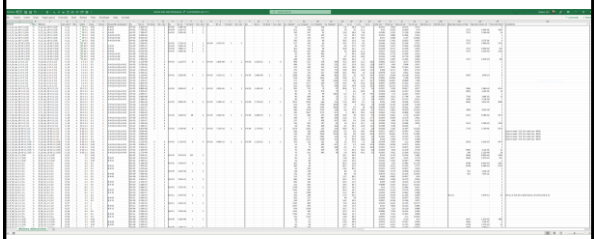
Lots and lots of data to look at

- Primary output is Excel table – used a custom tool by Tom Farris at NicheVision – extracted data from 6,119 runs
  6120 runs total
  +1 by hand
  - LR's of true donors for all mixtures have been captured
  - Largest non-donor LRs
  - Number of non-donors excluded (LR=0)
  - Template and variance values
- We will make this table available
- Searching and sorting on template or decon or conditioning options is interesting (At least I think so!)

25

25

It's hopeless to work through data on screen today



26

26

Exploratory data analysis

- I am not a statistician – I am a "practical-ician"
- I recruited an actual statistician
- James Curran is helping to look for data trends
- This is the current effort prior to submitting this data for publication

- For the most part, STRmix™ gives very, very similar LRs and database ranks throughout the study
- There are clear outliers scattered throughout

27

27

Relative error compared to cognate

- This looks at the difference between the "validated" STRmix™ kit and "error" range of the other seven kits
- This is the central 95% as we trimmed 2.5% from each end due to some extreme tails

Contributor	Mean relative error	Max relative error
K1	0.0423	11.3
K2	0.0499	0.936
K3	0.0626	0.593
K4	0.0824	0.331

$$\frac{1}{7}\sum_{i\neq\varphi}\frac{|\log_{10}(O_\varphi)-\log_{10}(O_i)|}{\log_{10}(O_\varphi)}$$

Obs Cognate Log(LR) – Obs Log(LR)
divided by
Obs Cognate Log(LR)

- K1 has most by far the most LRs, due to fully conditioned LRs

28

28

Example of an outlier

Kit	K1 LR	K2 LR
A	4.45E+20	1.09E+19
B	1.19E+20	2.80E+18
C	4.89E+21	1.16E+20
D	1.67E+19	4.14E+17
E	2.62E+23	6.30E+21
F	1.83E+17	4.52E+15
G	4.83E+20	1.17E+19
H	6.56E+15	2.80E+14

- Mixture B_14
- 3 person
- 3:2:1 ratio
- 1.0 ng input
- Decon #4 (cond'n on lowest template donor)

- K1 range (Rel. Error = 0.10)
  - 6.56E+15 6 Quadrillion
  - 2.62E+23 260 Septillion
- K2 range (Rel. Error = 0.22)
  - 2.80E+14 280 Trillion
  - 6.30E+21 6 Sextillion

29

29

Comparison of LR ranks for each kit

- Score each set of 8 LRs from lowest to highest LR
- Count them per STRmix™ kit

- Plot as histogram for each lab
  - 1 = lowest LR
  - 2 = 2nd lowest LR
  - …
  - 8 = largest LR

30

30

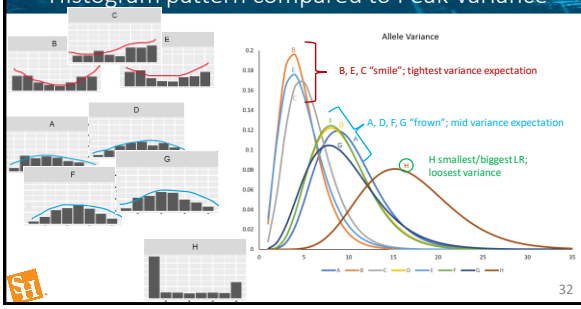
K1 LR rank histogram

- Lab H most often the smallest LR
- Seems to be a "smile" pattern
- Seems to be a "frowny" pattern

31

31

Histogram pattern compared to Peak Variance

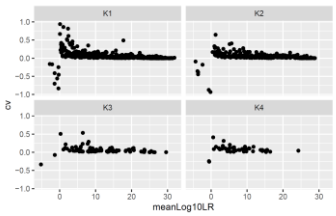


32

32

CV as a function of LR magnitude

- As LR gets larger, less variation across the kits
- As expected; less uncertainty for big LRs; generally more DNA

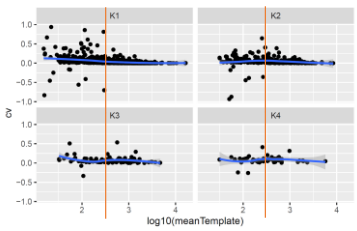


33

33

CV as a function of template

- By 315 rfu ($10^{2.5}$) the CV is <0.5
- Very good agreement between kits with peaks >300 rfu



34

34

Key Takeaways

- You get similar LRs using different STRmix™ parameters (within the one to two orders of magnitude)
- Interlaboratory STRmix™ parameters can effectively be used to detect ground truth donors from a mixture (similar LR rank)
  - We tested 10,000+ non-donor profiles
  - Continuing this work using 25 million profiles (DBLR™)
- Conditioning was the largest influence on the LR, not lab parameters
- STRmix™ seems to give "reliable" and "accurate" results regardless of the lab specific parameters used
- I suggest this shows STRmix™ is "trustworthy"

35

35

Thank you for listening!

- Thank you to Safia Boodoosingh for doing close to 10,000 STRmix™ runs overall with a repeat rate of <<1%
- Thank you to Hannah Kelly for serving as an external evaluator of Safia's project
- Thank you to James Curran for helping with the data analysis
- Thank you to ESR and the STRmix™ creators for supporting this project in various ways

36

36