MOVING TO PROBABILISTIC GENOTYPING

Todd W. Bille

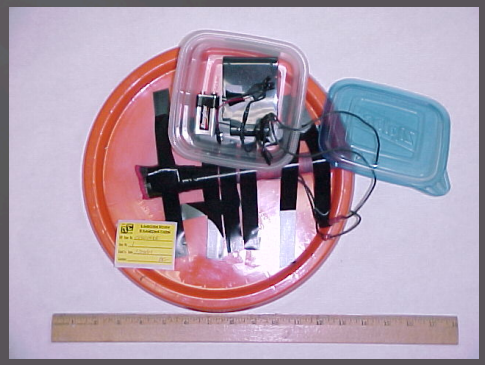
Bureau of Alcohol, Tobacco, Firearms and Explosives Laboratory

History of DNA at the ATF Laboratory

- DNA Unit created in May, 2006
- Initiated casework analysis in April, 2007
- 3 DNA analyst initially, now up to 4…will be hiring four more in the very near future

Concentrating on Touch Evidence

- Typical Evidence Examined at ATF Laboratory
  - Guns
  - Bomb components
  - Molotov Cocktails
- >90% of our evidence samples are from "touch evidence"



Initial Methods - General

- Cotton swabs
- Double swab technique (wet/dry)
- Qiagen QiaAmp Micro DNA extraction
- Microcon 100 concentration
- Identifiler
- 3130 Genetic Analyzer
- AT: 50 rfu, ST: 200 rfu
- CPI for mixtures

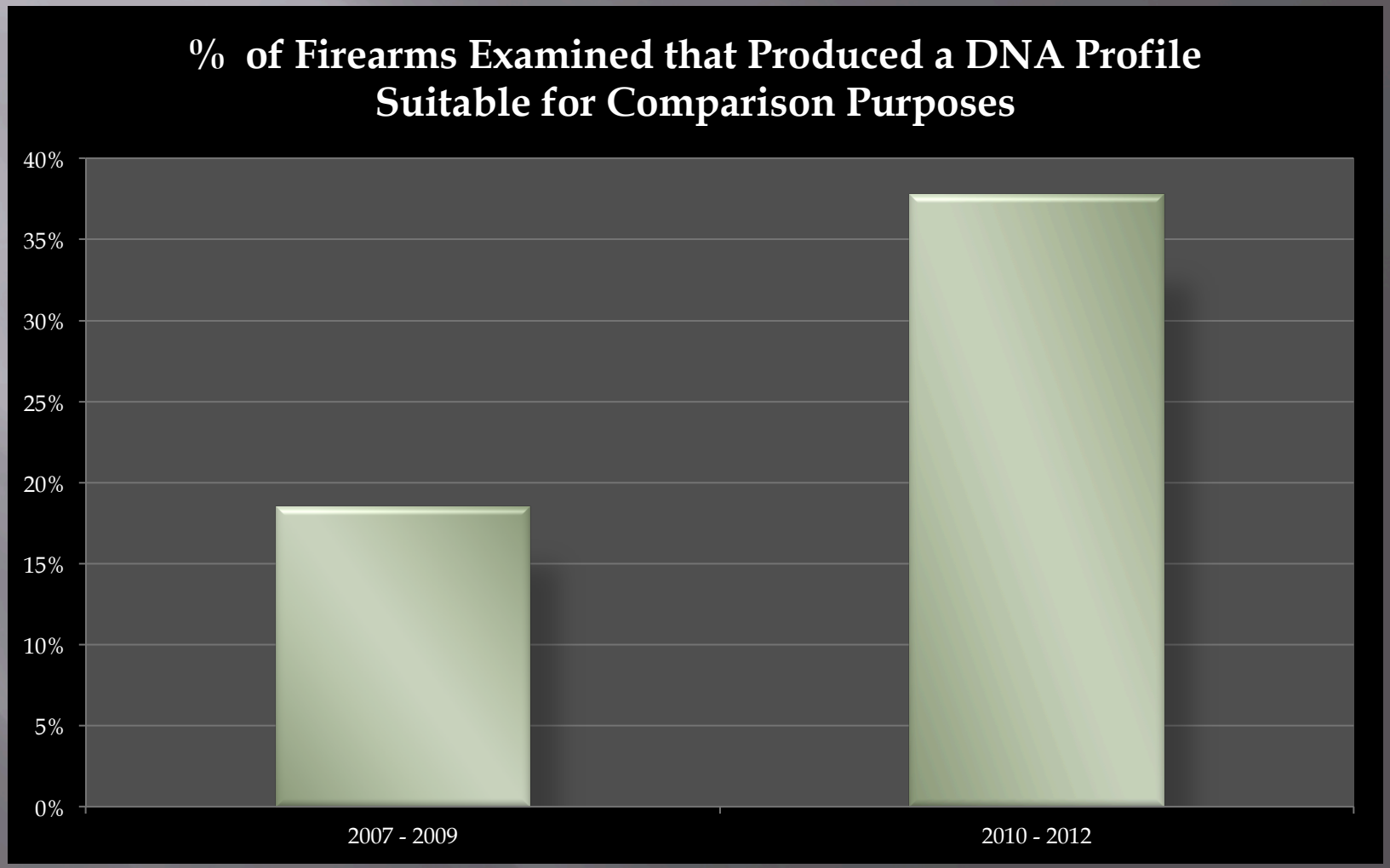
Process Optimization – Things to Consider

- Collection
  - Type of swab
  - Fluid used
  - Collection method – double swab (wet/dry)
- Extraction
  - What extraction method
  - Optimizing the extraction method
  - Maximize *amplifiable* DNA
- Concentration
  - Minimize the loss of DNA during the concentration step
- Interpretation / Statistical Approach

Changes to Methods

- Swab textured areas before latent print processing then after dye-staining
- 2 X AL buffer modification to Qiagen QiaAmp Micro DNA extraction
- QiaAmp Investigator Kit DNA extraction
- Minimize amount of swab material placed in lysis buffer
- Microcon 30
- Carrier RNA used during Microcon
- Incorporation of 2p rule - 2011
- RMP for mixtures - 2011

Comparison of First Three Years to Last Three Years



Suitable for Comparison Purposes

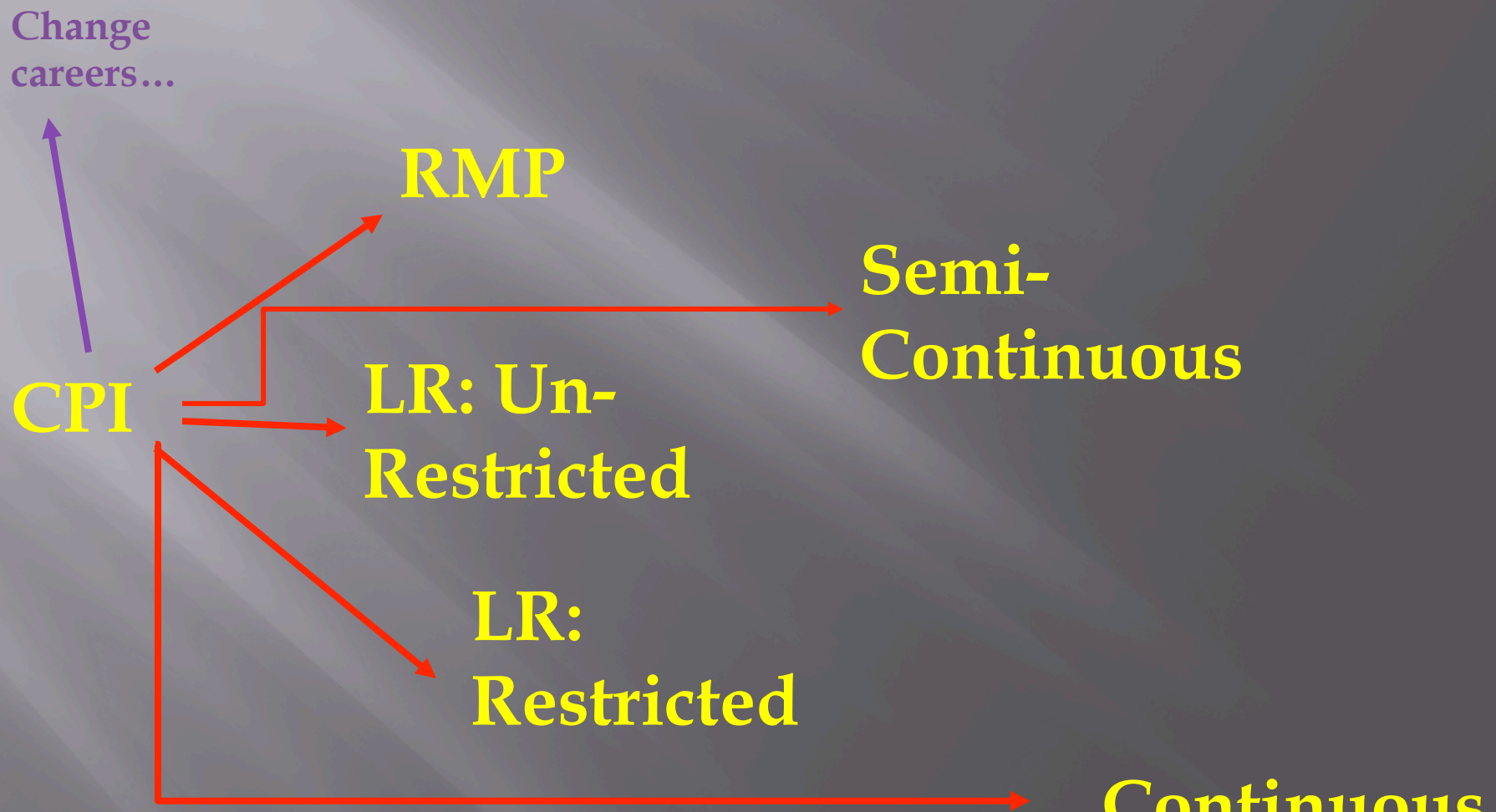
Initial Methods

- CPI Statistical Approach
  - Stochastic Threshold of 200 rfu
  - If a locus had one or more alleles below the stochastic threshold, the locus could not be used for statistical purposes
  - This was true for single source samples, as well
  - As a result, much data was being discarded

Initial Methods

- CPI Statistical Approach
  - DNA profiles containing no peaks above the stochastic threshold , even assumed single source profiles, could be used for exclusionary purposes only.
  - Issues arise when suspect A excluded, suspect B can neither be included or excluded.

Needed to make a change



Application of Random Match Probability Calculations to Mixed STR Profiles

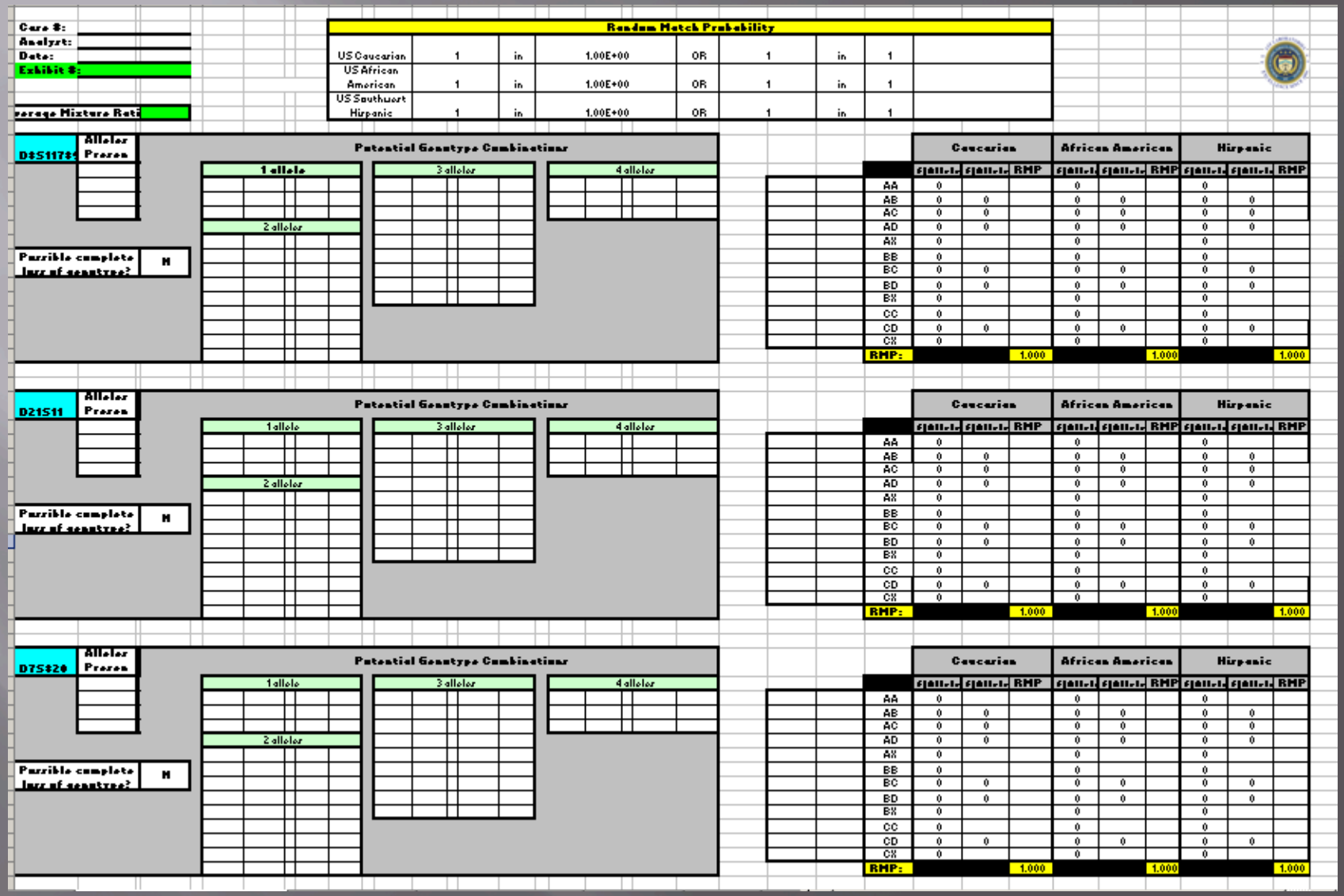
Todd Bille M.Sc., Jo-Anne Bright M.Sc. and John Buckleton Ph.D.

Journal of Forensic Sciences, Vol. 58, Issue 2, pages 474–485, March 2013

Random Match Probability (RMP)

- Makes better use of the observed data than CPI and unrestricted LR, but still does not use reference profile data like an LR
- Still a binary approach
- Interpretation documented prior to comparison to known samples
- Significant amount of time for interpretation
- Variation between analysts
- Outliers cause issues…

Documentation



Case #:
Analyst:
Date:
Exhibit #:

Average Mixture Ratio

Random Match Probability

US Caucasian	1	in	1.00E+00	OR	1	in	1
US African American	1	in	1.00E+00	OR	1	in	1
US Southwest Hispanic	1	in	1.00E+00	OR	1	in	1

D3S1358

Alleles Present:

Possible complete # of genotypes?: H

Potential Genotype Combinations

1 allele | 2 alleles | 3 alleles | 4 alleles

	Caucasian			African American			Hispanic		
	f(all.1)	f(all.1)	RMP	f(all.1)	f(all.1)	RMP	f(all.1)	f(all.1)	RMP
AA	0			0			0		
AB	0	0		0	0		0	0	
AC	0	0		0	0		0	0	
AD	0	0		0	0		0	0	
AX	0			0			0		
BB	0			0			0		
BC	0	0		0	0		0	0	
BD	0	0		0	0		0	0	
BX	0			0			0		
CC	0			0			0		
CD	0	0		0	0		0	0	
CX	0			0			0		
RMP:			1.000			1.000			1.000

D21S11

Alleles Present:

Possible complete # of genotypes?: H

Potential Genotype Combinations

1 allele | 2 alleles | 3 alleles | 4 alleles

	Caucasian			African American			Hispanic		
	f(all.1)	f(all.1)	RMP	f(all.1)	f(all.1)	RMP	f(all.1)	f(all.1)	RMP
AA	0			0			0		
AB	0	0		0	0		0	0	
AC	0	0		0	0		0	0	
AD	0	0		0	0		0	0	
AX	0			0			0		
BB	0			0			0		
BC	0	0		0	0		0	0	
BD	0	0		0	0		0	0	
BX	0			0			0		
CC	0			0			0		
CD	0	0		0	0		0	0	
CX	0			0			0		
RMP:			1.000			1.000			1.000

D7S820

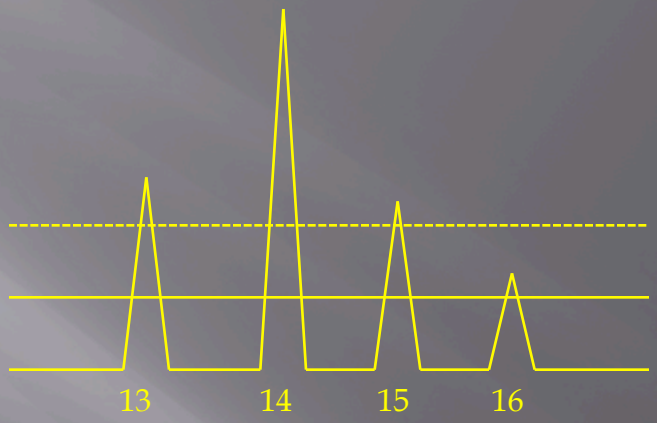
Alleles Present:

Possible complete # of genotypes?: H

Potential Genotype Combinations

1 allele | 2 alleles | 3 alleles | 4 alleles

	Caucasian			African American			Hispanic		
	f(all.1)	f(all.1)	RMP	f(all.1)	f(all.1)	RMP	f(all.1)	f(all.1)	RMP
AA	0			0			0		
AB	0	0		0	0		0	0	
AC	0	0		0	0		0	0	
AD	0	0		0	0		0	0	
AX	0			0			0		
BB	0			0			0		
BC	0	0		0	0		0	0	
BD	0	0		0	0		0	0	
BX	0			0			0		
CC	0			0			0		
CD	0	0		0	0		0	0	
CX	0			0			0		
RMP:			1.000			1.000			1.000



D8S11789	Alleles Present
	13
	14
	15
	16

Possible complete loss of genotype?	N

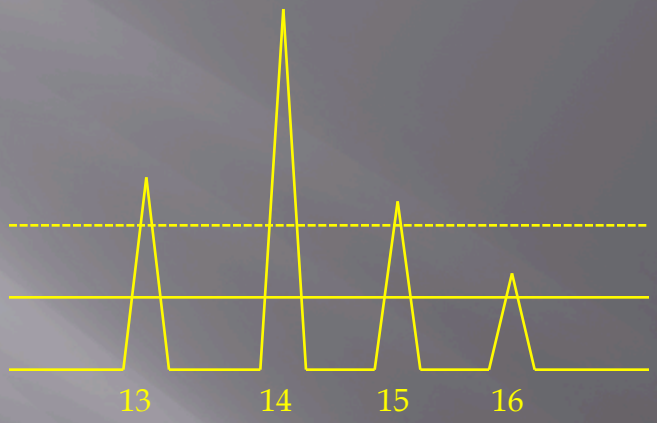
Potential Genotype Combinations

1 allele				

2 alleles				

3 alleles				

4 alleles				
13	14	y	15	16
13	15	n	14	16
13	16	y	14	15



			Caucasian			African American			Hispanic		
			f(Allele 1)	f(Allele 2)	RMP	f(Allele 1)	f(Allele 2)	RMP	f(Allele 1)	f(Allele 2)	RMP
		AA									
13	14	AB	0.339	0.202	0.137	0.222	0.333	0.148	0.325	0.246	0.160
		AC									
13	16	AD	0.339	0.013	0.009	0.222	0.044	0.020	0.325	0.025	0.016
		AX									
		BB									
14	15	BC	0.202	0.110	0.044	0.333	0.214	0.143	0.246	0.116	0.057
		BD									
		BX									
		CC									
15	16	CD	0.110	0.013	0.003	0.214	0.044	0.019	0.116	0.025	0.006
		CX									
		RMP:			0.192			0.329			0.239

Binary Approach Issues

Probability of Drop-out Pr(D)

- The probability that drop-out occurred associated with peaks below the stochastic threshold is not equal across the range of peak heights.

Probability of Drop-out Pr(D)

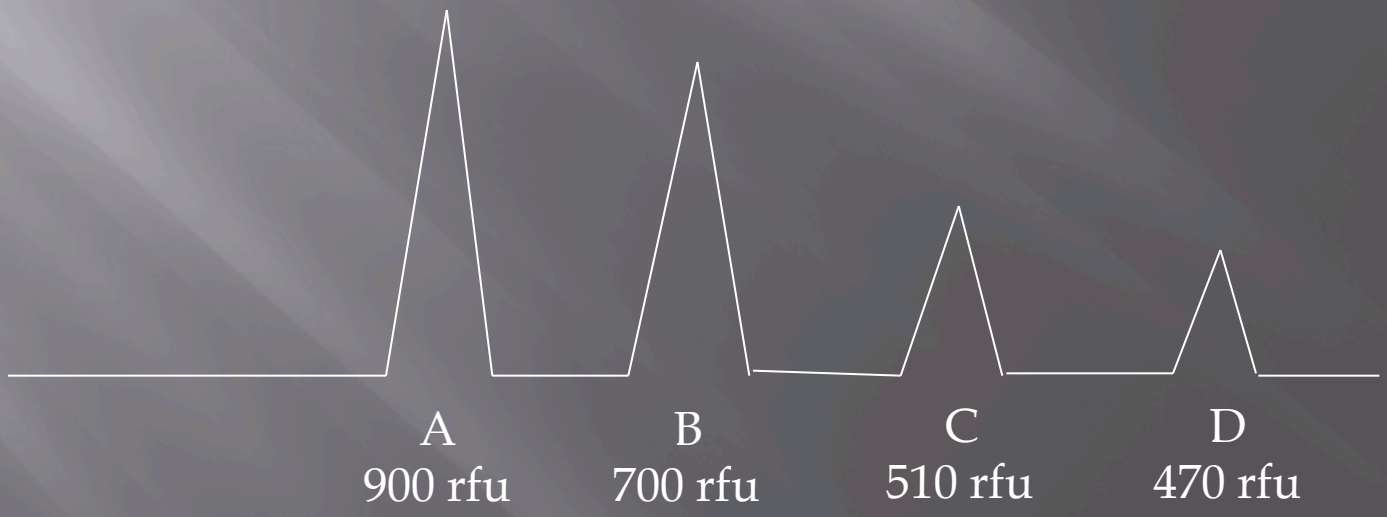
- Therefore, it is not always conservative to use the *2p rule* or drop the locus for statistical purposes.
- If the suspect is an A,B at the locus below, the Pr(D) approaches zero as the allelic peak nears the Stochastic Threshold and the *2p rule* becomes less conservative.

Peak Height Ratio Expectations

- Similar to drop-out, binary methods do not differentiate between genotype combinations that may fall within expectations, but would be unlikely.

Peak Height Ratio Expectations

- At the locus below, the A peak at 900 rfu can be expected to have a sister peak as low as 504 rfu. The B peak can have a sister peak as low as 355 rfu. Therefore, based on these expectations, the A,C / B,D combination ***does*** meet the peak height ratio expectations. However, it is much more likely that the mixture is composed of an A,B and C,D.

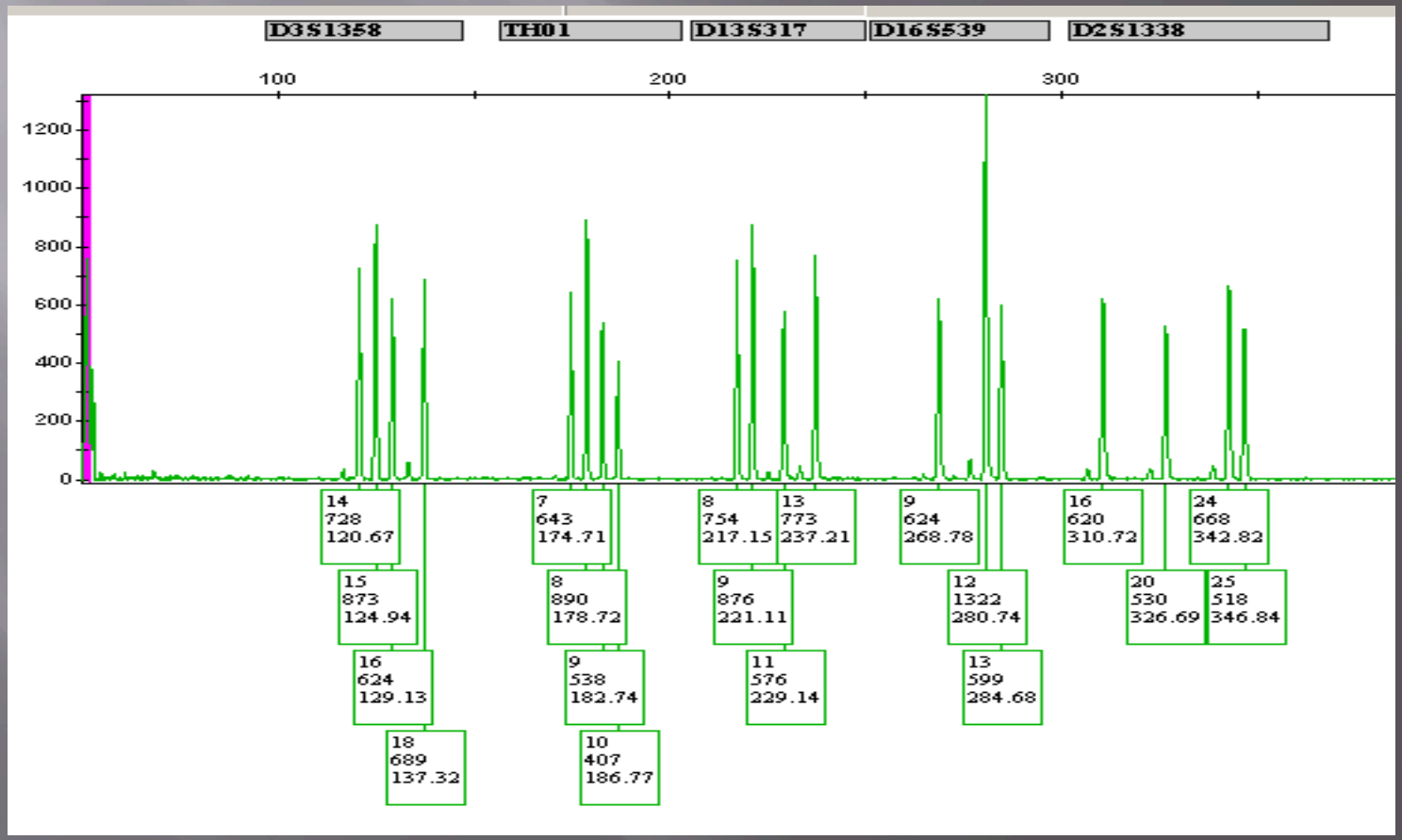


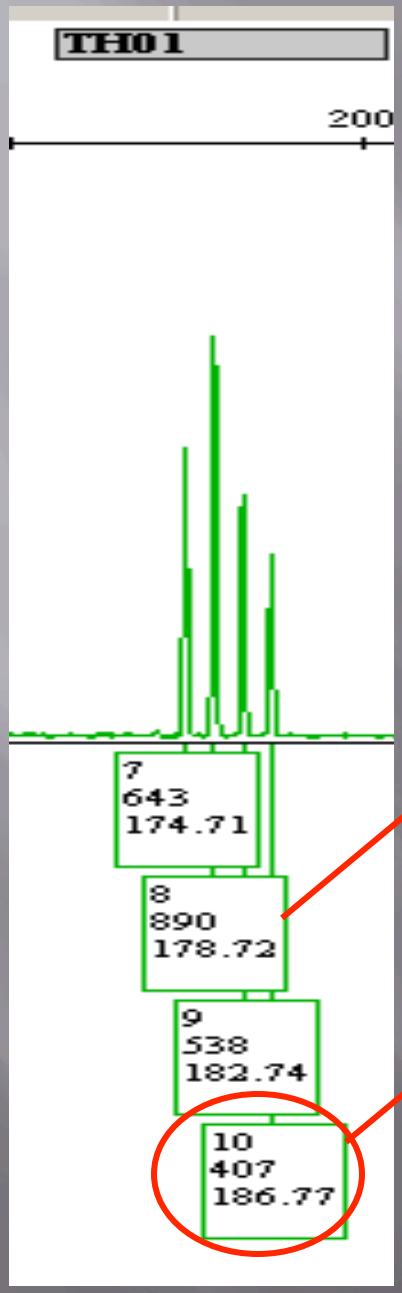
Binary Methods – Results outside expectations

- The ranges of the various parameters (stutter, mixture ratio, peak height ratio, etc.) cannot account for every incident.
- Events may occur outside expected ranges due to primer mutation, tri-allele loci, degradation, significant differences in allele size within a locus, etc.
- These events may occur randomly, as well. Ranges are typically defined by +/- three standard deviations. This range therefore should capture 99.7% of the events. This means 0.3% (or about 1 out of 300) events will fall outside the range.
- The analyst must be aware of this during interpretation.

Binary Methods – Results outside expectations

1 : 1 Mixture, 500 pg total DNA template





It would appear the 7 and 8 pair and the 9 and 10 pair.

Expected range goes as low as 497 rfu.

8,10 Genotype Excluded

This is a known mixture of an 8,10 and 7,9

Time of Transition

CPI

Probabilistic Approach

RMP

Desired aspects of continuous model

- Relatively user-friendly
- Published formulae, not a “black box”
- Made efficient use of data, e.g. peak height balance, probability of drop-out, stutter, mix ratio
- Ability to customize using data from our lab
- “Consistency”

Desired aspects of continuous model

- Cost
- Computer requirements
- Speed of analysis
- Ability to perform calculations by hand
- Ability to obtain an output file of calculations

STRmix™

- Relatively easy user interface
- Reasonable cost
- Can be run on a laptop or desktop computer
- Typical analysis times are 1 – 3 minutes, longer for more complex mixtures
- Supporting publications
- All formulae used are available
- Can produce output files showing all calculations if desired (very large file)

STRmix™

- Continuous Model
- Calculates an LR
- Uses Markov Chain Monte Carlo method (MCMC), therefore a different result will be calculated using the exact same data. However, results will be relatively close and the range can be defined during validation.
- Considers the following:
  - Degradation curve
  - Per allele stutter ratios
  - Drop-in
  - Drop-out
  - Peak variance
  - Locus-specific amplification efficiencies

STRmix™

- Genotype combinations that best fit the observed data given greater weight
- All genotype combinations considered, less probable combinations (outliers) are given less weight

STRmix™

- Models used in STRmix™ are developed from laboratory's data
- Model Maker
  - Approximately 90 samples
  - Varying quantities and qualities
  - Should be done for each kit

Current Status

- All analysts have attended the one week STRmix™ training
- Software has been purchased and recently received
- Will begin the validation process this summer
- Awaiting input from various groups as to exactly what needs to be validated for this type of software

Thank you for your attention

Questions?