25th Congress of the
International Society for Forensic Genetics
Melbourne, Australia

An Investigation of Software Programs Using "Drop-out" and "Continuous" Methods for Complex Mixture Interpretation

Michael D. Coble and John M. Butler

National Institute of Standards and Technology

September 6, 2013

NIST and NIJ Disclaimer

<u>Funding</u>: Interagency Agreement between the National Institute of Justice and NIST Office of Law Enforcement Standards

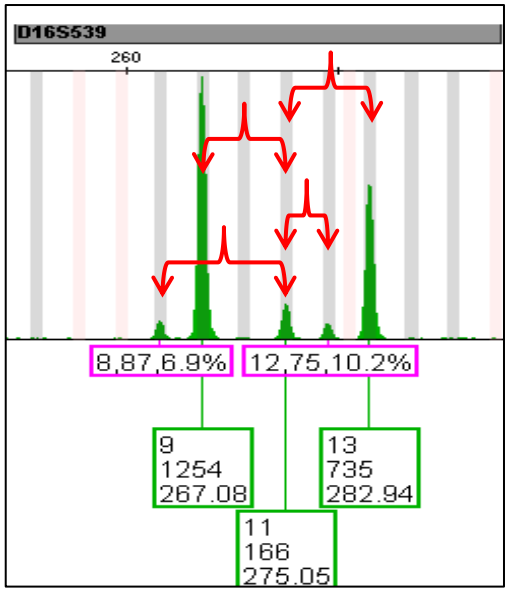
Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

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.

Challenging Mixtures - Uncertainty

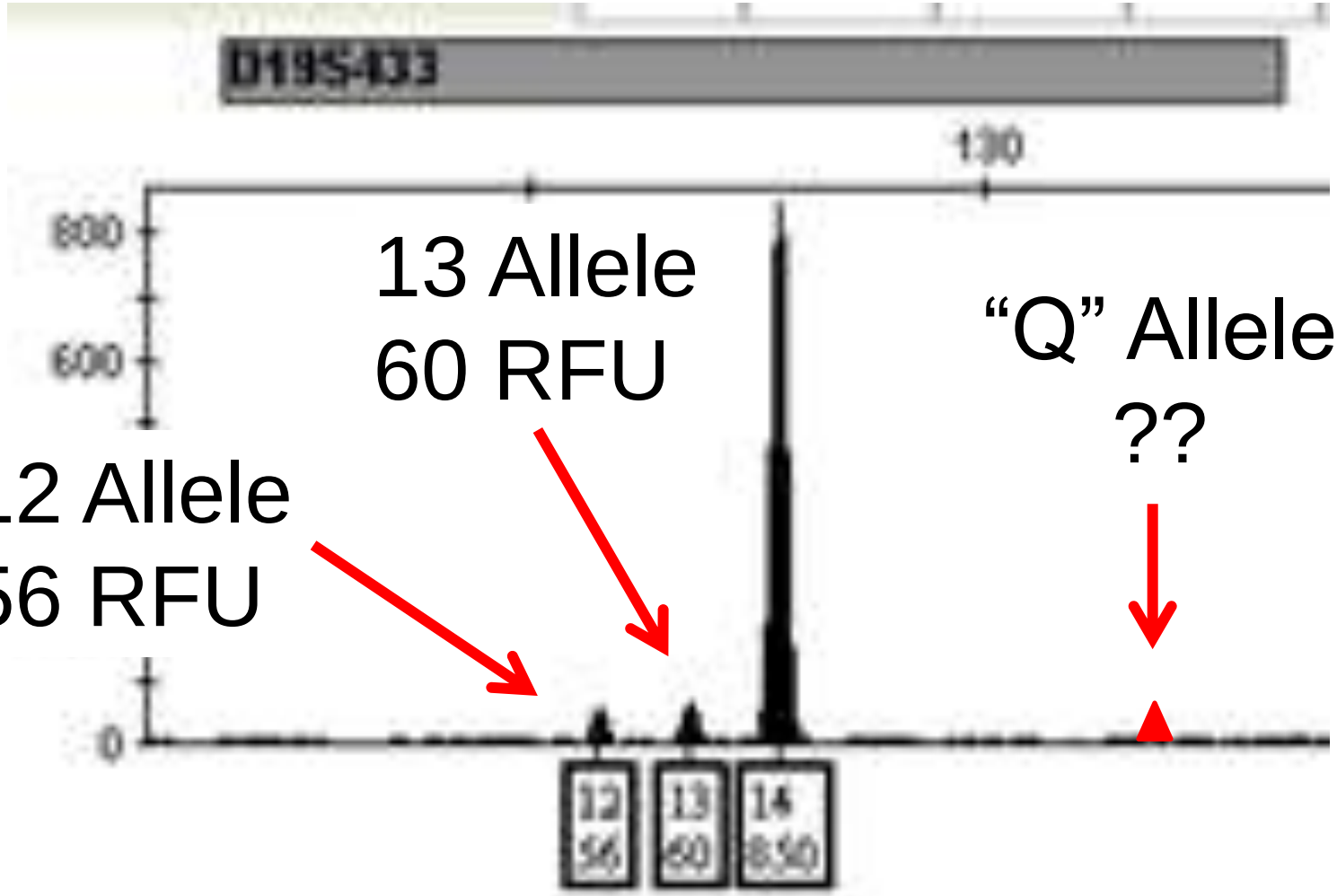
- **If allele dropout is a possibility** (e.g., in a partial profile), then there is uncertainty in whether or not an allele is present in the sample…and therefore what genotype combinations are possible

- **If different allele combinations are possible** in a mixture, then there is uncertainty in the genotype combinations that are possible…



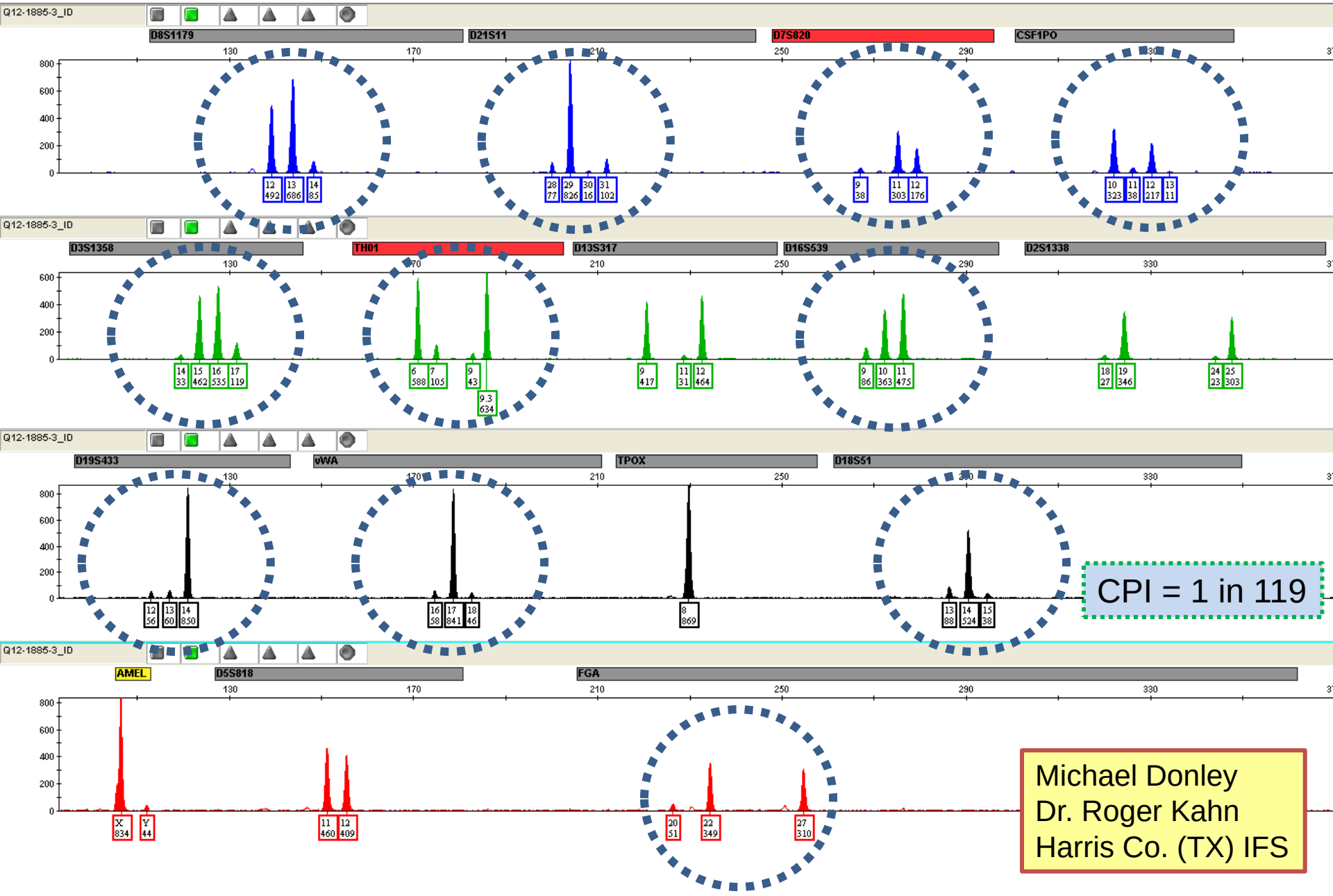
Possible allele pairing with the 11

Challenging Mixtures



How to handle low level data

- Continue to use RMNE (CPI, CPE)

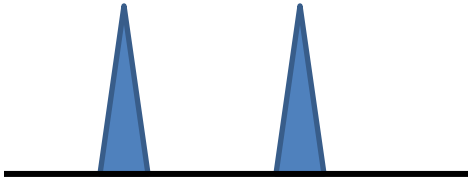


Michael Donley
Dr. Roger Kahn
Harris Co. (TX) IFS

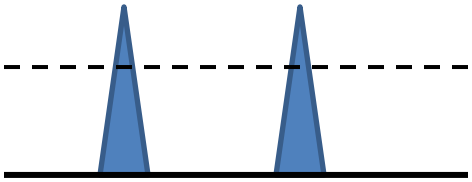
How to handle low level data

- Continue to use RMNE (CPI, CPE) (not optimal)
- Use the Binary LR with 2p

Suspect

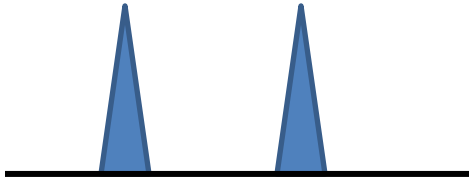


Evidence

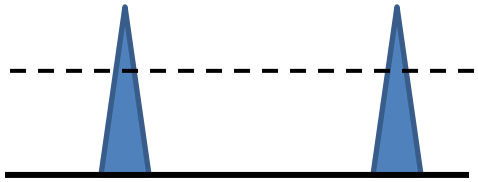


$$LR = \frac{1}{2pq}$$

Suspect

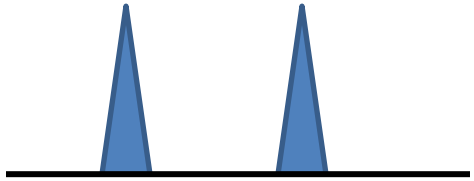


Evidence

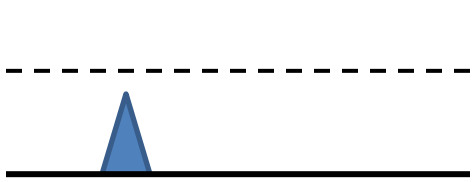


$$LR = \frac{0}{2pq}$$

Suspect



Evidence



$$LR = \frac{?}{2pq}$$

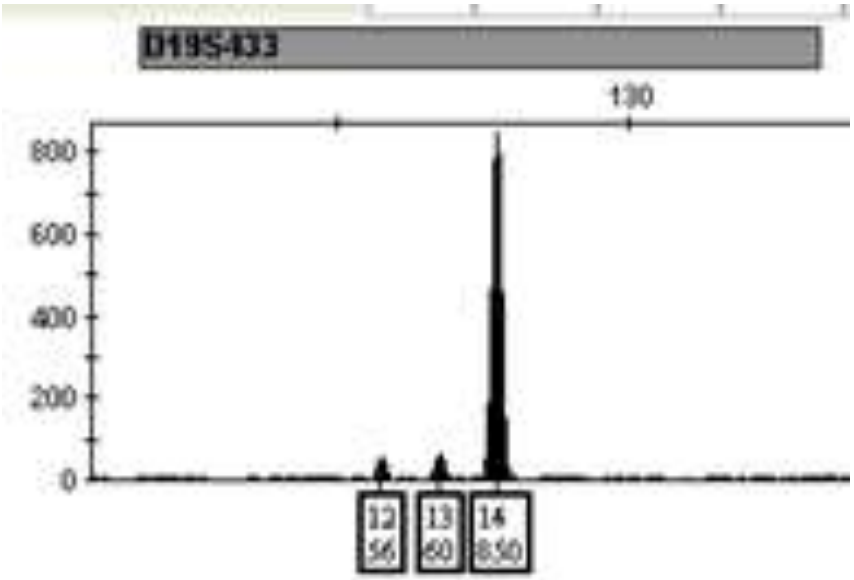
The Binary LR approach

“2p”

How to handle low level data

- Continue to use RMNE (CPI, CPE) (not optimal)
- Use the Binary LR with 2p (not optimal)
- Semi-continuous methods with a LR (Drop models)

$Pr(D_{out}) = 0.89$



Some Drop Model Examples

- LR mix (Haned and Gill)
- LikeLTD (Balding and Buckleton)
- Lab Retriever (Lohmueller, Rudin and Inman)
- FST (NYOCME, Mitchell *et al.*)
- Kelly *et al.* (University of Auckland, ESR)
- Puch-Solis *et al.* (LikeLiRa and LikeLiRaHT)

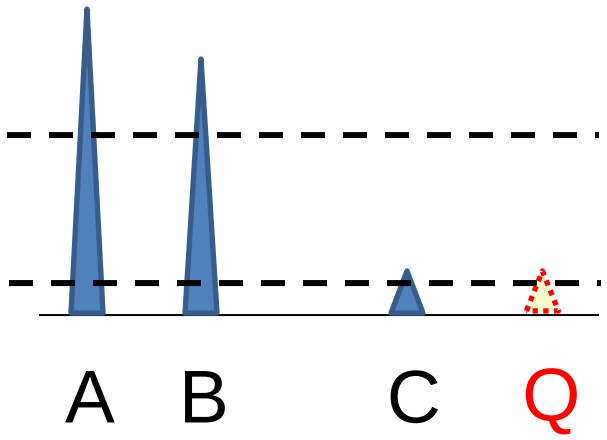
The drop models only use the alleles present in the mixture

How to handle low level data

- Continue to use RMNE (CPI, CPE) (not optimal)
- Use the Binary LR with 2p (not optimal)
- Semi-continuous methods with a LR (Drop models).
- Fully continuous methods with LR.

Continuous Models

- Mathematical modeling of “molecular biology” of the profile (mix ratio, PHR (Hb), stutter, etc…) to find optimal genotypes, giving **WEIGHT** to the results.



Probable Genotypes

AC – 40%

BC – 25%

CC – 20%

CQ – 15%

Some Continuous Model Examples

- TrueAllele (Cybergenetics)
- STRmix (ESR and Australian collaboration)
- Cowell et al. (FSI-G (2011) **5:**202-209)

Weights are determined by performing simulations of the data (Markov Chain Monte Carlo - MCMC)

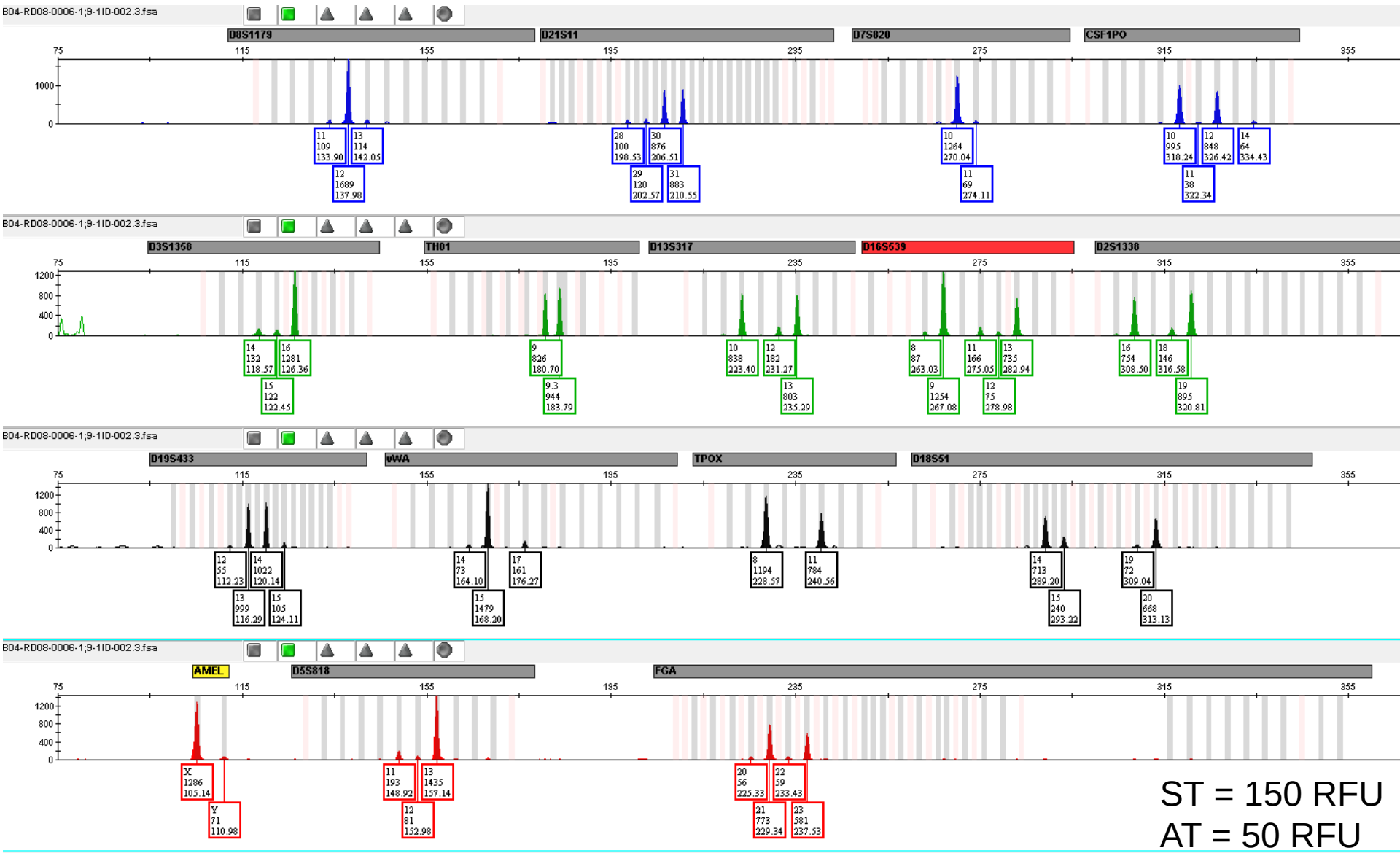
Software Examined

- LR Mix (Gill and Haned) – open source R program with GUI.
- Lab Retriever (Rudin, Lohmueller, Inman) – free software, based on the Balding/Buckleton approach.
- TrueAllele (Cybergenetics) – Continuous approach, publications and presentations by Perlin *et al.*
- STRmix (Australia/NZ) – Continuous approach, publications by Taylor, Bright and Buckleton.

Some Ground Rules

- For LRmix and Lab Retriever, the same values for $Pr(D_{out})$ and $Pr(D_{in})$ were used.
- The NIST (2003) allele frequencies for Western Europeans were used for all systems.
- TrueAllele analyses were performed at both 10 RFU (default) and 50 (2p mixtures) or 30 (3p mixture) RFUs.

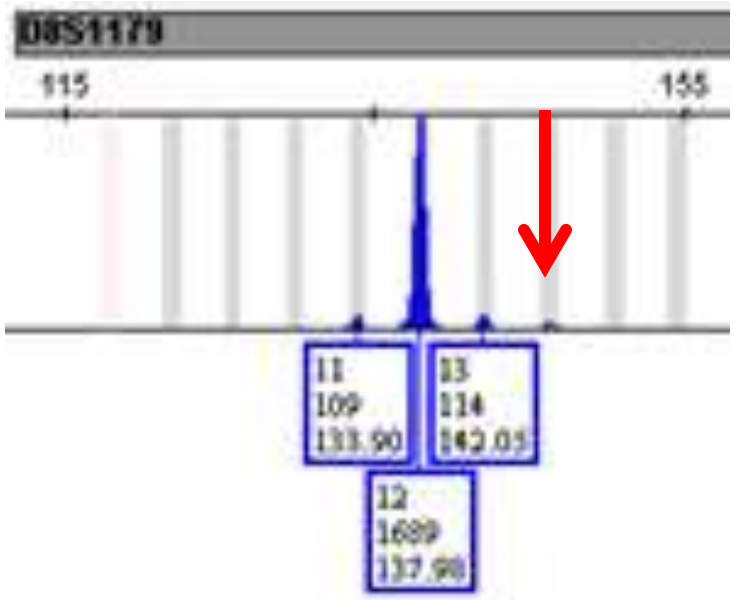
Example 1 – Low-level 2p with D_{out}



ST = 150 RFU

AT = 50 RFU

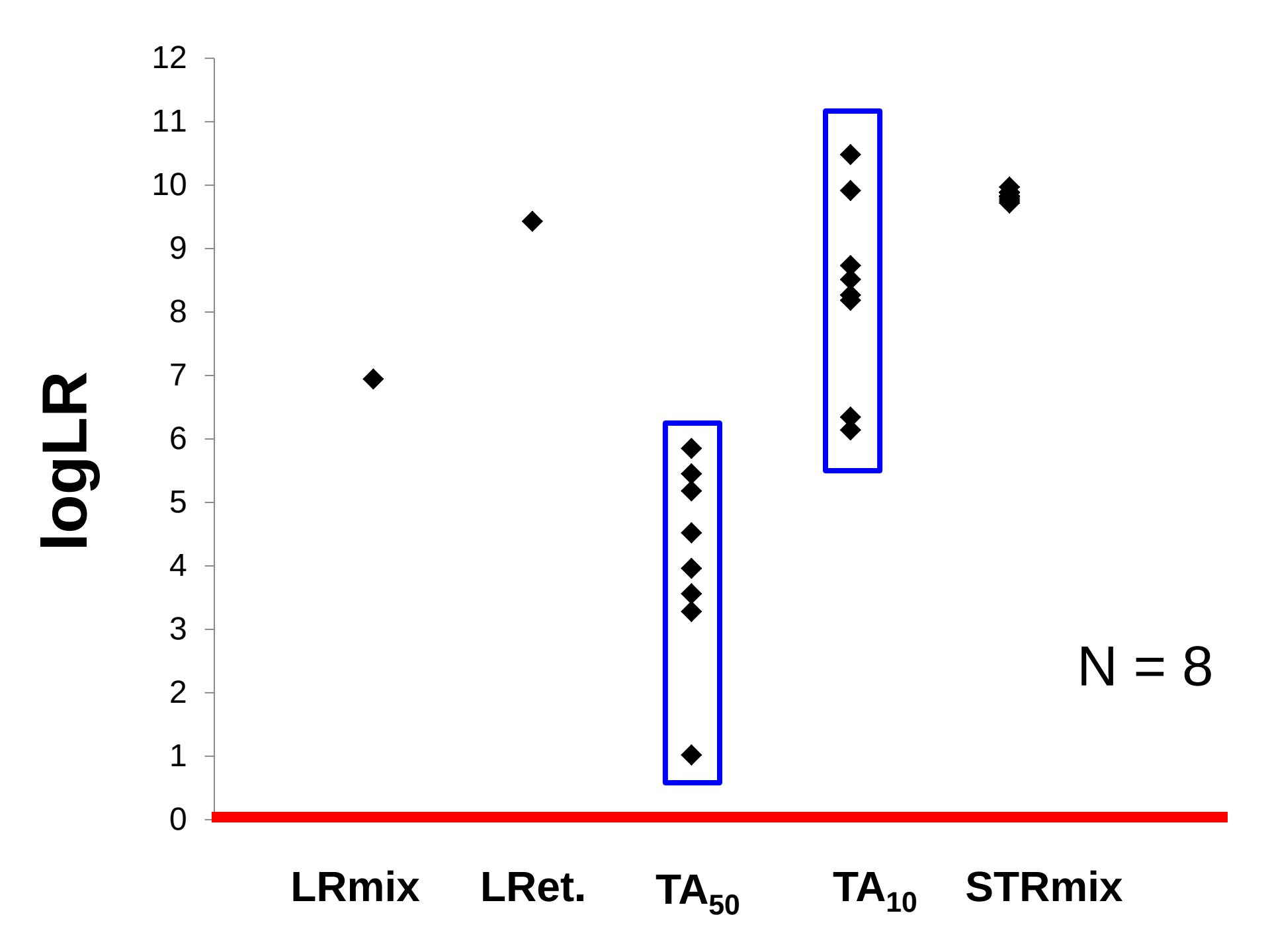
Loci with Drop-out



POI = 13, 14

POI = 21, 24

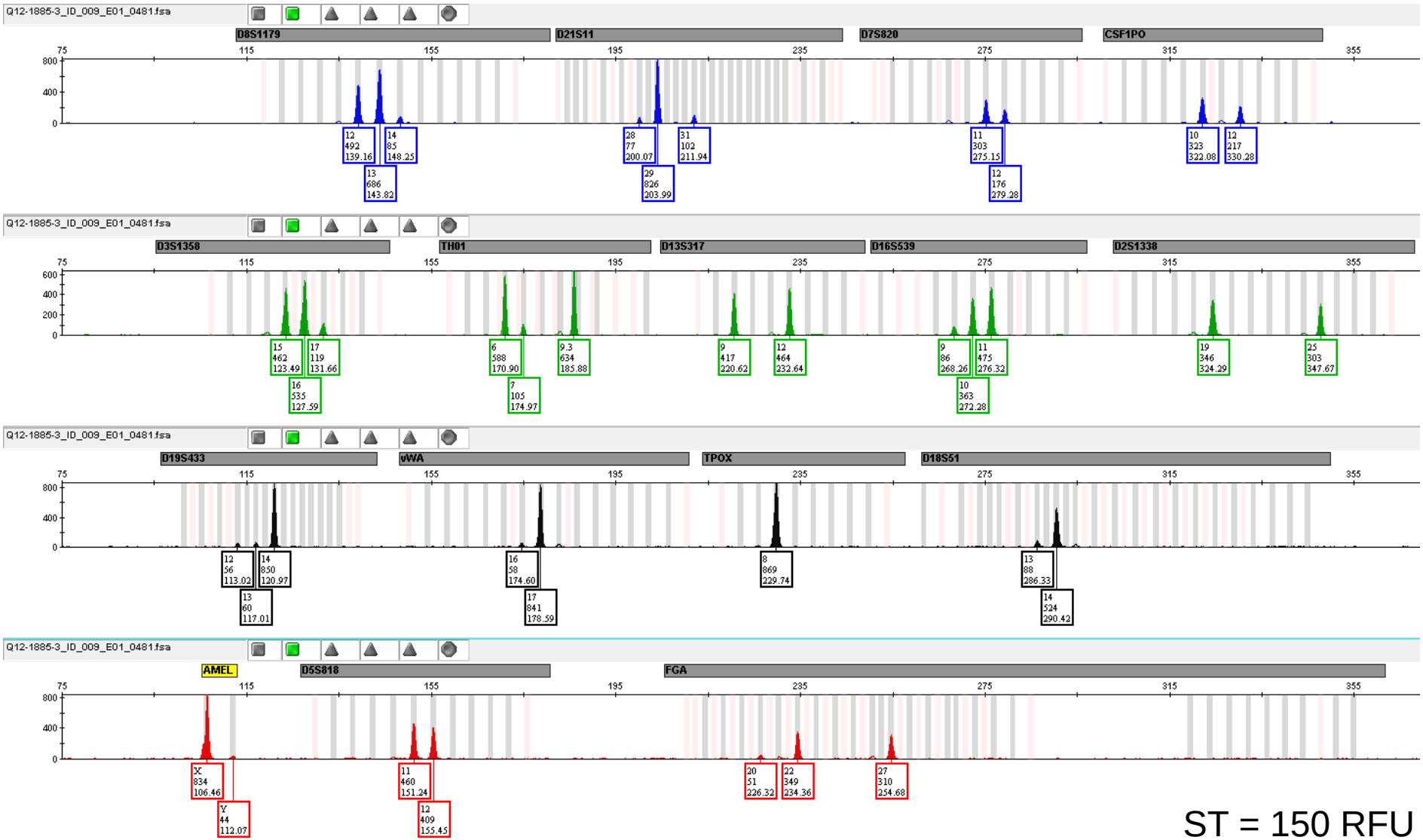
2 loci with allele drop-out


logLR
12
11
10
9
8
7
6
5
4
3
2
1
0
N = 8
LRmix
LRet.
TA50
TA10
STRmix

Time of Analysis

	2p - 2DO
LRmix	< 1 sec.
Lab Ret.	< 1 sec.
TrueAllele	8-12 hours
STRmix	25.2 sec.

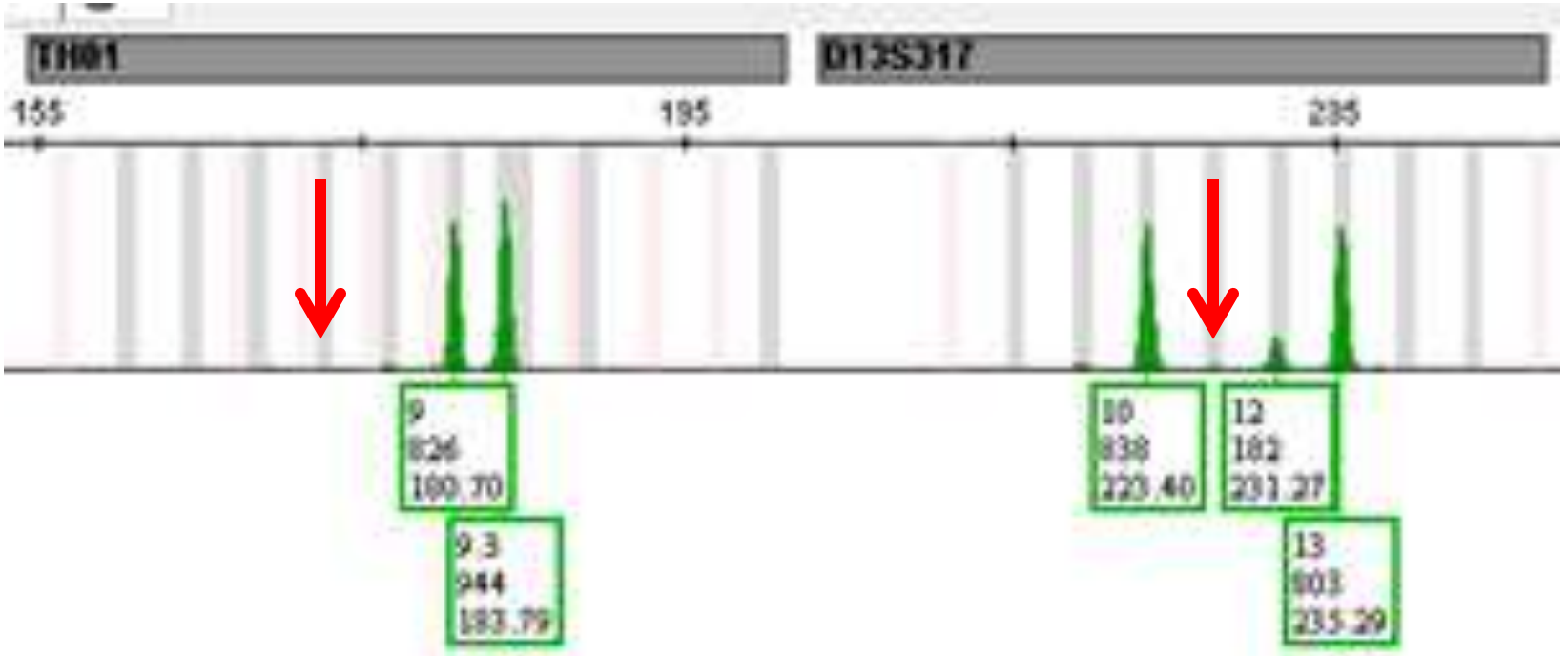
Example 2 – Low-level 2p more D_{out}



ST = 150 RFU

AT = 50 RFU

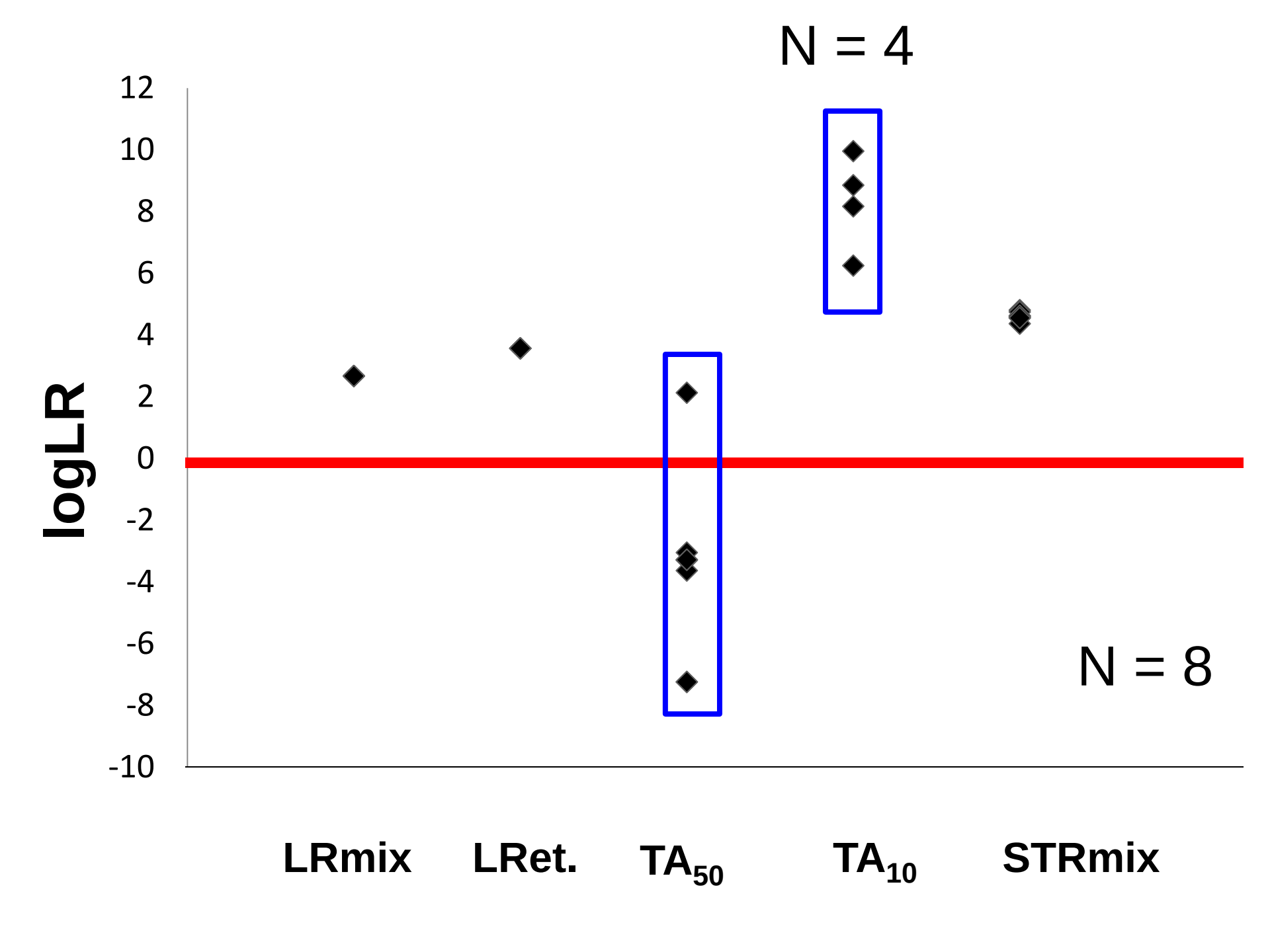
Loci with Drop-out



POI = 7, 9

POI = 11, 12

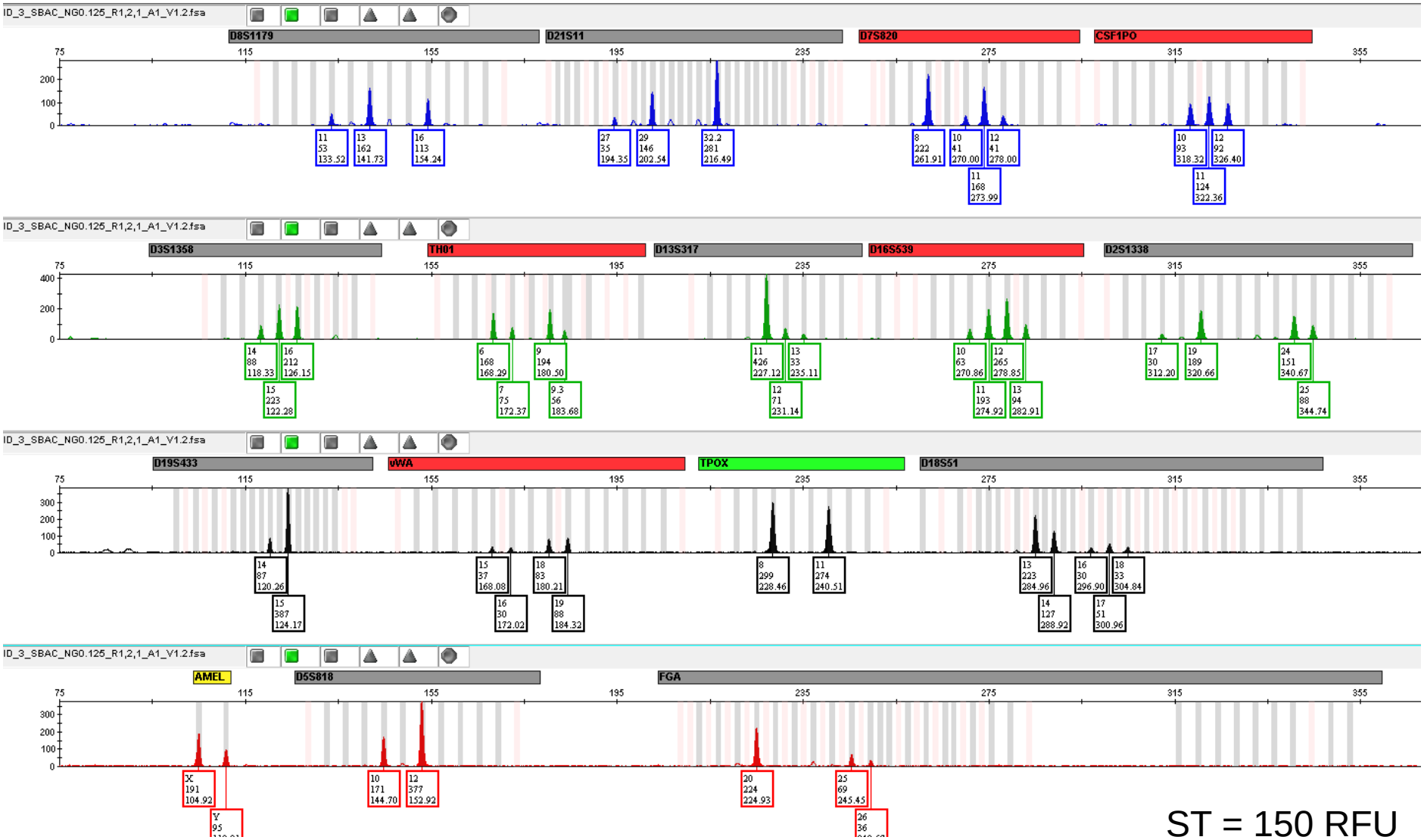
5 loci with allele drop-out; 1 locus drop-out (CSF)


N = 4
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
logLR
N = 8
LRmix
LRet.
TA50
TA10
STRmix

Time of Analysis

	2p - 2DO	2p - 6DO
LRmix	< 1 sec.	< 1 sec.
Lab Ret.	< 1 sec.	1 sec.
TrueAllele	8-12 hours	8-12 hours
STRmix	25.2 sec.	14.8 sec.

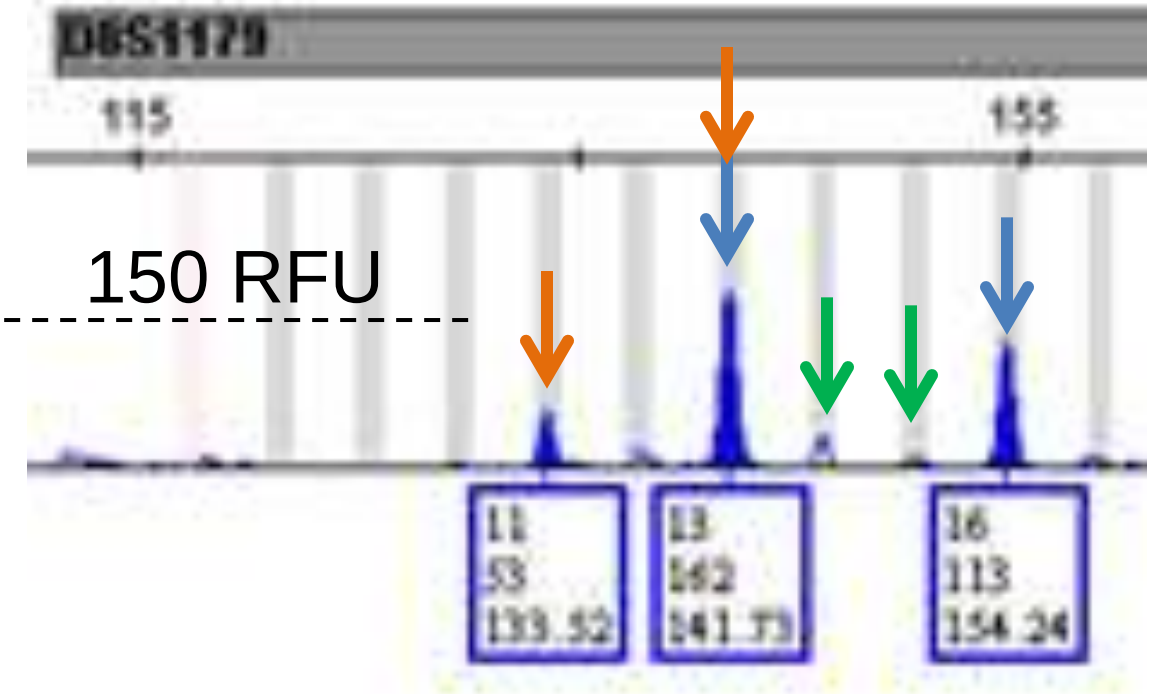
Example 3 – Low-level 3 person mixture



ST = 150 RFU

AT = 30 RFU

Example 3 – Low-level 3 person mixture



A = 13,16

B = 11,13

C = 14,15

125 pg input DNA

1:2:1 ratio

B = female

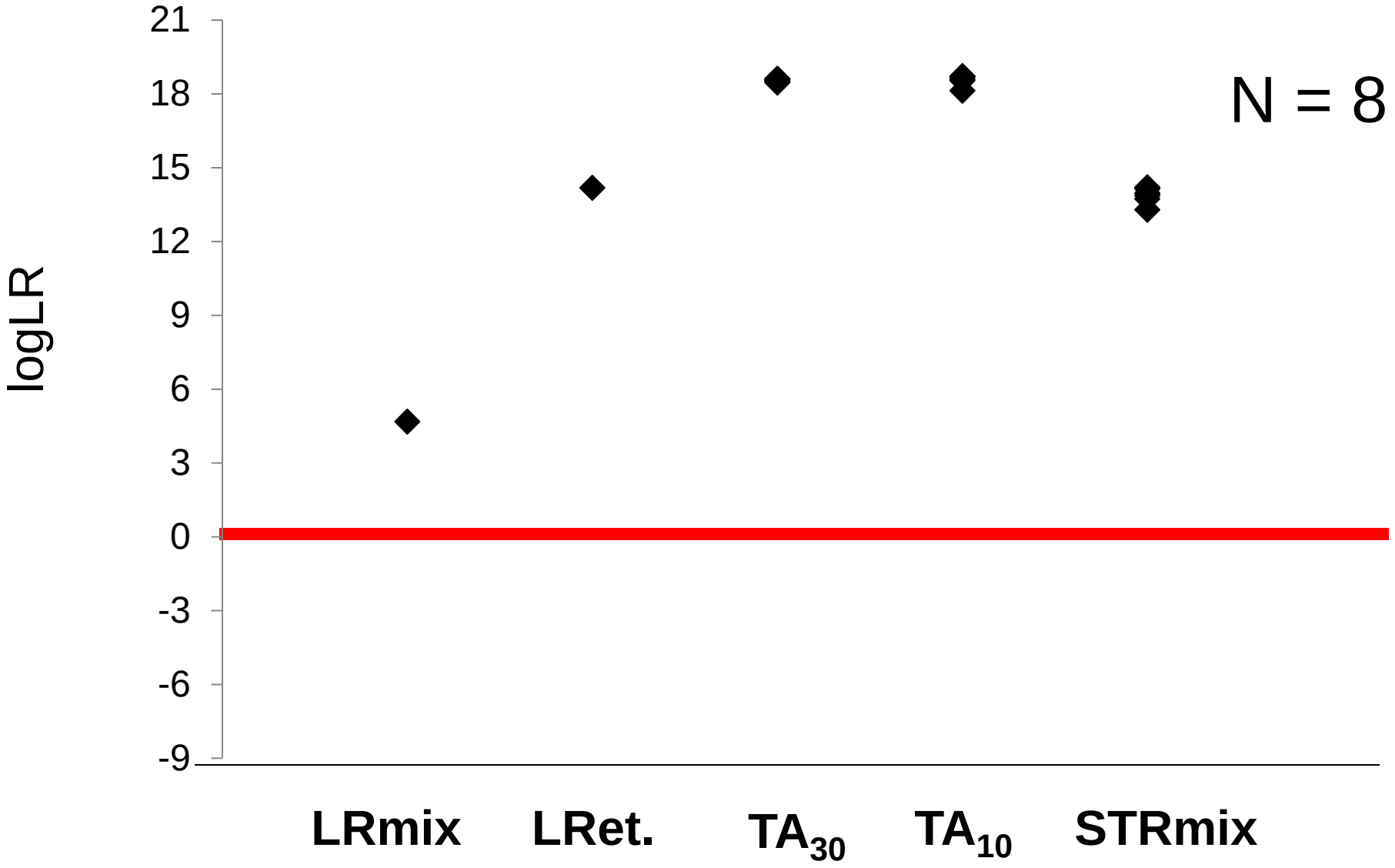
Conditioning

- H_P = B (vic) and C (suspect 1) and A (suspect 2)
- (1) H_D = B (vic) and C (suspect 1) and 1 Unk
- (2) H_D = B (vic) and A (suspect 1) and 1 Unk

Suspect A, Pr(DO) = 0.02

Suspect C, Pr(DO) = 0.529

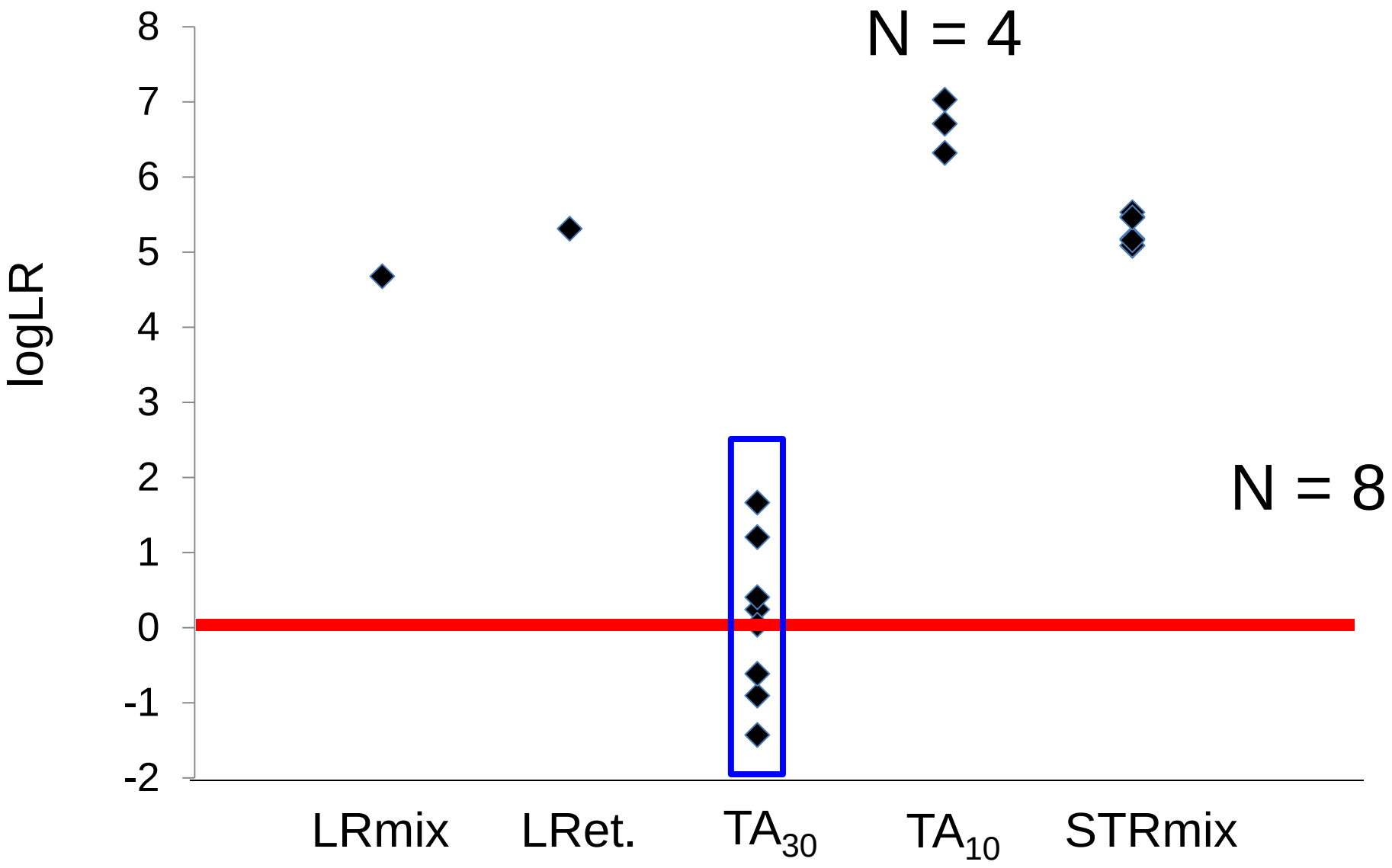
H_D = B (vic) and C (suspect 1) and 1 Unk (A)



Time of Analysis

	2p - 2DO	2p - 6DO	3p - A unk
LRmix	< 1 sec.	< 1 sec.	3 sec.
Lab Ret.	< 1 sec.	1 sec.	8 sec.
TrueAllele	8-12 hours	8-12 hours	16 hours +
STRmix	25.2 sec.	14.8 sec.	63.1 sec.

H_D = B (vic) and A (suspect 1) and 1 Unk (C)



Time of Analysis

	2p - 2DO	2p - 6DO	3p - A unk	3p - C unk
LRmix	< 1 sec.	< 1 sec.	3 sec.	3.4 sec.
Lab Ret.	< 1 sec.	1 sec.	8 sec.	7.5 sec.
TrueAllele	8-12 hours	8-12 hours	16 hours +	16 hours +
STRmix	25.2 sec.	14.8 sec.	63.1 sec.	50.5 sec.

Summary

- Probabilistic Methods make better use of the data than RMNE or the binary LR with 2p.
- The goal of the software programs should not be to simply “get bigger numbers” but to understand the details of these approaches and not treat the software as a “black box.”
- Semi-continuous approaches will produce a LR that could be replicated by hand if necessary.

Summary

- Each approach has its own advantages and disadvantages.

- “When analysed using a discrete model such as that of [Balding and Buckleton], it is necessary to rely on an analyst designations of low peaks as allelic, stutter, or masking and hence ambiguous.”

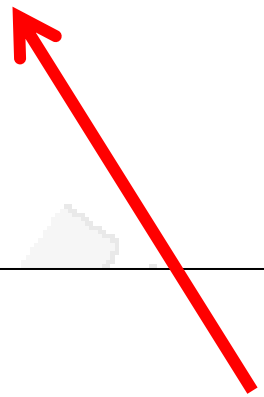
  - Puch-Solis *et al.* (2013)

Future Studies

- Examination of other probabilistic programs.
- Include challenging four person mixtures.
- Determining the risk of including a suspect not in the mixture using randomly generated profiles.

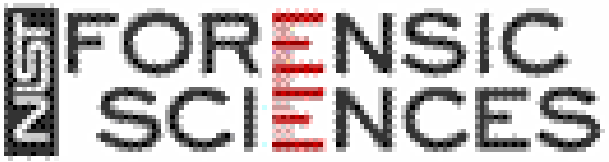
NIST MIX13 Interlaboratory Study

- DNA Advisory Board Quality Assurance Standards
- Interlaboratory Studies
- NIST Mixture 2005 Interlab Study MIX05 Data
- **NIST Mixture 2013 Interlab Study MIX13 Data** NEW
- Validation information
- DNA Quantitation - SRM 2372
- Technology for resolving STR alleles



http://www.cstl.nist.gov/strbase/interlab/MIX13.htm

NIST Interlaboratory Mixture Interpretation Study 2013 (MIX13)



Study Details

Five cases are provided each with an evidentiary sample file (a mixture of at least one suspe profile, suspect(s) profiles and other known references. We have generated .fsa files on an . (BU)/IdentifilerPlus (NIST) kits. Allelic ladders, positive and negative controls are also pro

Case #1 – represents sexual assault evidence from the sperm fraction of a vaginal swab. Th comparison. Data available in Identifiler Plus and PP16HS.

Case #2 – represents evidence swabbed from the handle of a handgun retrieved outside the from four gang members are provided for comparison. Data generated by BU in either Iden Identifiler sample – all other examples used the GS500 LIZ size standard.

Case Information

Reference Profiles for Comparison Purposes

(Excel file with different suspects for each case in individual tabs)

Identifiler Plus Data

Zip file containing all samples from the 5 cases

Case 1

Evidence	Ladder	Positive control	Negative control

Case 2

Evidence	Ladder	Positive control	Negative control

PowerPlex 16 HS Data

Zip file containing all samples from the 5 cases

Acknowledgments

National Institute of Justice and NIST OLES

Applied Genetics Group
Dr. Peter Vallone
Becky Hill
Erica Butts

Dr. Charlotte Word
Dr. Robin Cotton
Dr. Catherine Grgicak

Dr. Roger Kahn
Michael Donley

Contact info:
mcoble@nist.gov
+1-301-975-4330