Understanding "Null" Alleles and STR Allele Mobility Issues through Variant Allele Sequencing

Margaret C. Kline and John M. Butler

American Academy of Forensic Sciences 2008 Meeting
February 23, 2008 Washington D.C.

Disclaimers

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and 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.

Our publications and presentations are made available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

Subjects to be Covered

- SNPs
  - Causing allele dropout
  - Not causing allele dropout
- Deletions
- Nomenclature
  - Why a repeat motif is selected
- Alleles outside of the "normal" range
- Tri-Allelic Samples
  - "normal" tri-allelic pattern
  - Alleles outside the "normal" range.

Loci We Currently Sequence

- We have sequencing primers for the CODIS loci, including Penta D & E, D2S1338, and D19S433.
- We also have sequencing primers for many of the Y-STR loci.

http://www.cstl.nist.gov/biotech/strbase

Lab Resources and Tools

- Addresses for scientists working with STRs
- Training Materials
- STR Allele Sequencing

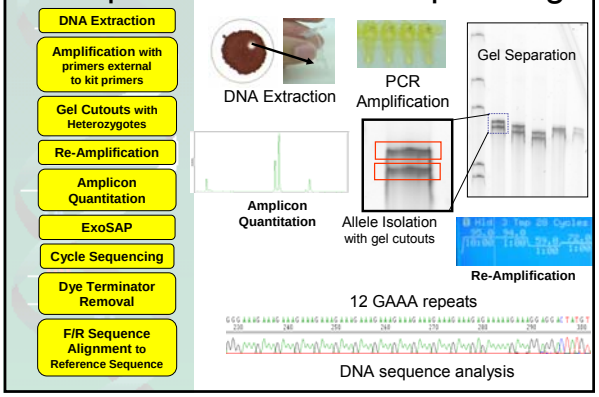
STRBase has a summary of alleles that have been submitted and sequenced, **if** the submitting agency agrees to share the information.

We require a minimum of 10 ng for the sequencing.
We request copies of the electropherograms demonstrating the variant allele.
The more information we have up front the better.
Please have patience we will get to your samples!

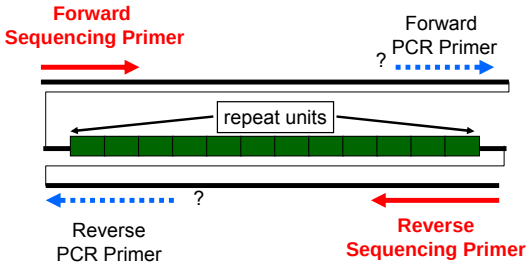
Sample Submissions

- For those that desire more assurances of confidentiality we can have MOUs signed.
- We generally re-type the samples at NIST prior to starting sequencing.
- We may run a monoplex assay (single locus).
- We return results as PowerPoint slides.
- We thank all of those agencies that have used this free service (thanks to NIJ)!

Steps in STR Allele Sequencing

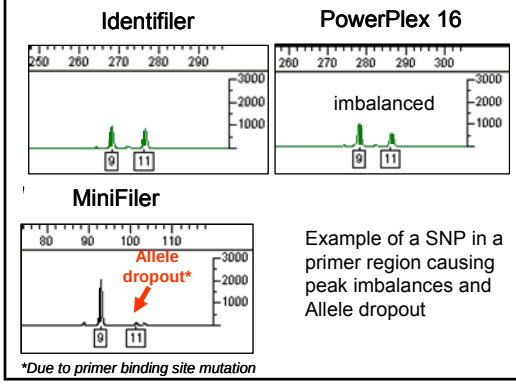


Sequencing Primers



Sequencing primers are designed to encompass the normal PCR primer locations based on PCR product sizes, or published primer sets.

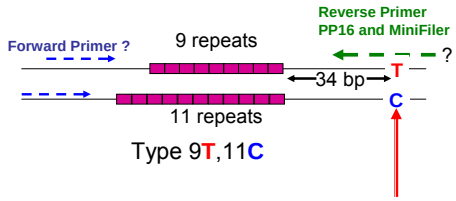
D16S539 SRM 2391b Genomic 8



Example of a SNP in a primer region causing peak imbalances and Allele dropout

Due to primer binding site mutation

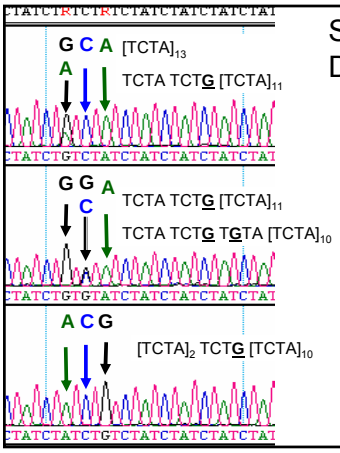
D16S539 SRM 2391b Genomic 8



This T→C mutation 34 bp downstream of the repeats causes allelic dropout!

Locus Sequence Variability

- In collaboration with Tom Hall (IBIS) has analyzed some of the NIST population samples by Mass Spec methods.
- In many samples the Mass Spec detected SNPs, base pair changes within specific loci.
- We have gone back and sequenced some of these samples to determine where the SNPs were located.



SNPs within the D8S1179 repeat

Repeat is TCTA
Three NIST samples have genotypes 13,13.

Analysis by Mass Spec indicates the presence of SNPs (Tom Hall, IBIS) Confirmation of the Mass Spec by sequencing at NIST indicates:

There are **4** different 13 alleles in these 3 samples.

Base Pair difference between Repeats

D8S1179

$[TCTA]_{13}$
TCTA TCT**G** $[TCTA]_{11}$
TCTA TCT**G** T**G**TA $[TCTA]_{10}$
$[TCTA]_2$ TCT**G** $[TCTA]_{10}$

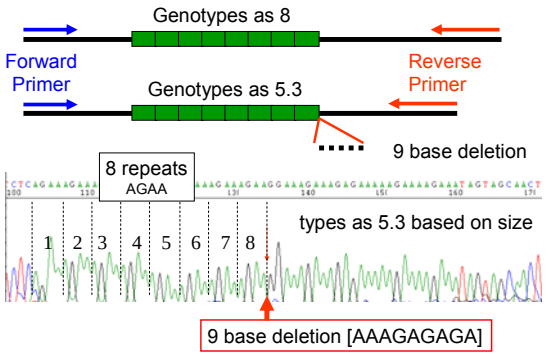
Allele	AVG	SD	N	
8	123.82	0.02	10	
				4.1
9	127.90	0.02	7	
				4.1
10	132.03	0.06	42	
				4.1
11	136.17	0.05	35	
				4.3
12	140.42	0.07	80	
				4.5
13	144.93	0.08	153	
				4.4
14	149.36	0.07	131	
				4.3
15	153.67	0.07	89	
				4.3
16	157.93	0.07	21	
				4.1
17	162.02	0.04	2	

D16S539

Mass Spec detected fewer SNPs in this locus.
Less average bp variability seen between repeat sizes.

Allele	AVG	SD	N	
5	252.08	0.01	2	
				4.0
8	264.09	0.11	17	
				4.0
9	268.14	0.11	93	
				4.1
10	272.20	0.12	56	
				4.0
11	276.18	0.14	162	
				4.1
12	280.25	0.14	158	
				4.1
13	284.30	0.12	72	
				4.1
14	288.35	0.07	8	

D18S51 deletion results in 5.3 Allele



DYS712 results

Repeats	Repeat Motif	bp size
23	$[AGAT]_{18}$ $[AGAC]_5$	177.49
22.3	$[AGAT]_2$ GAT $[AGAT]_{14}$ $[AGAC]_6$	176.42

There is a loss of a base **A** in the third AGAT repeat.
There is also difference in the number of AGAC repeats

Determining Repeat Motifs

- Sequencing results must agree with Genotyping results.
- Sometimes this comparison results in an initial simple repeat motif becoming more complex.
- A number of samples must be typed and sequenced for these determinations to be made.

Determining Repeat Motifs DYS715

ID	Allele	Repeat Motif	Size (bp)
A	14	$[TAGA]_{14}$	195.4
B	11	$[TAGA]_{11}$	183.5
C	12	$[TAGA]_{12}$	187.4
D	13	$[TAGA]_{13}$	191.3
E	12?	$[TAGA]_{12}$	191.4

The genotyping data does not match the sequencing data using a simple Repeat Motif of TAGA.

Determining Repeat Motifs DYS715

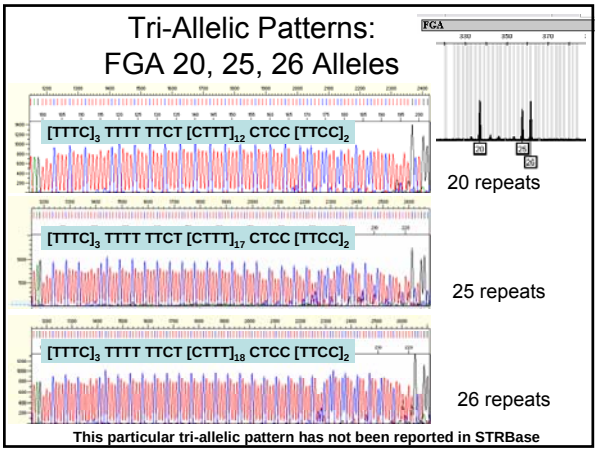
ID	Allele	Repeat Motif	Size (bp)
A	24	$[TAGA]_{14}$ N_{20} $[TGGA]_{10}$	195.4
B	21	$[TAGA]_{11}N_{20}$ $[TGGA]_{10}$	183.5
C	22	$[TAGA]_{12}$ N_{20} $[TGGA]_{10}$	187.4
D	23	$[TAGA]_{13}$ N_{20} $[TGGA]_{10}$	191.3
E	23	$[TAGA]_{12}$ N_{20} $[TGGA]_{11}$	191.4

With the complex repeat motif, the genotyping and the sequencing results agree!

Penta D 10, Variant Allele 19

All sequenced bases align before and after the repeat region. The 19 allele has been previously reported in STRBase. The Penta D ladder has Alleles 2.2, 3.2, 5, 7 – 17 represented.

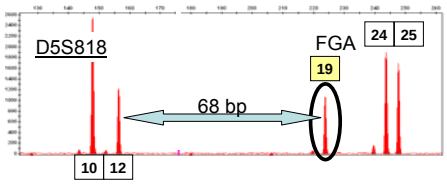
Tri-Allelic Patterns: FGA 20, 25, 26 Alleles



This particular tri-allelic pattern has not been reported in STRBase

Is this an FGA - Tri-allelic pattern identified using Identifiler?

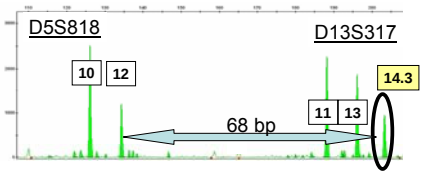
PK HT Ratio
12/10 - 0.48



PK HT Ratio
19/24 - 0.55
25/24 – 0.89

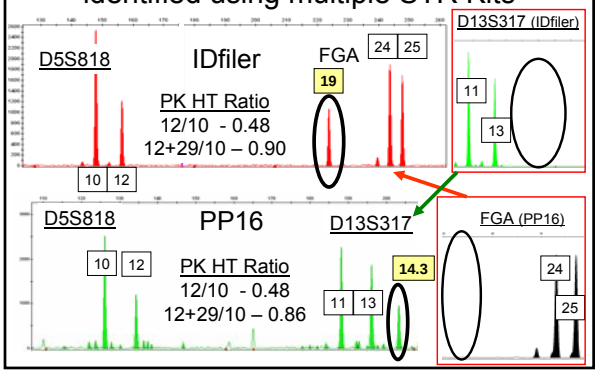
Or is this a D13S317 - Tri-allelic pattern identified using Powerplex 16?

PK HT Ratio
12/10 - 0.48

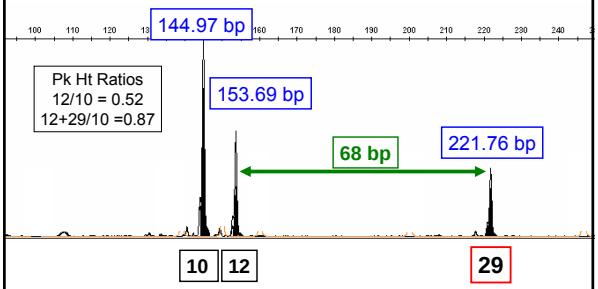


PK HT Ratio
13/11 - 0.83
14.3/11 – 0.42

It's really a D5S818 Tri-allelic pattern identified using multiple STR Kits

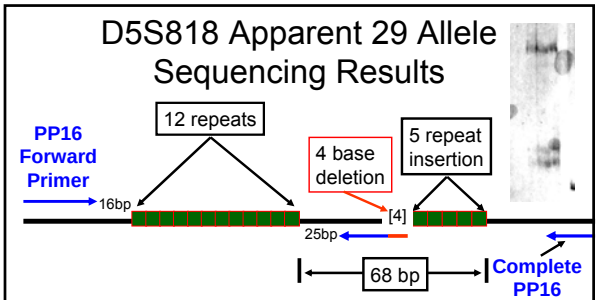


D5S818 monoplex results



The 68 bp size difference between the 12 allele and the variant allele sizing as an **"apparent 29"** allele.

D5S818 Apparent 29 Allele Sequencing Results



There is a 4 bp deletion, the last 4 bases of the PP16 reverse primer binding site, followed by an insertion of 5 repeats.
The 10 and 12 alleles of this sample have been sequenced and have the expected sequences.

Are there other large D5S818 alleles?

- STRBase Tri-allelic reports for FGA for 19,*,* patterns with AB amplification kits.
  - 5 reports :
  - 19,20,21; 19,20,23; 19,20,24; 19,22,23; 19,24,25
  - But there we have sequenced true tri-allelic FGA samples
- STRBase Tri-allelic reports for D13S317 for *,*, OL patterns with PP16 amplification kits.
  - NO tri-allelic patterns with Off-Ladder alleles reported

Thanks to Sample Contributors

- Some of the Laboratories that have contributed samples for sequence analysis include:
  - Maryland State Police
  - Armed Forces DNA Identification Laboratory
  - FSS and Kuwait government lab
  - Nebraska State Crime Lab
  - DNA Solutions
  - Peter de Knijff's lab at Leiden University
  - Westchester County Forensic Science Laboratory
  - UNTHSC DNA Identity Laboratory
  - Harris County Medical Examiner
  - The Commonwealth of Massachusetts Dept of State Police
  - ATF
  - DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines Diliman

Thank you for your Attention!!

Acknowledgements

Funding:

Interagency Agreement between National Institute of Justice and NIST Office of Law Enforcement Standards

NIST Project Team:

John Butler, Pete Vallone, Margaret Kline, Jan Redman, Amy Decker, Becky Hill, Dave Duewer

Email: Margaret.Kline@nist.gov