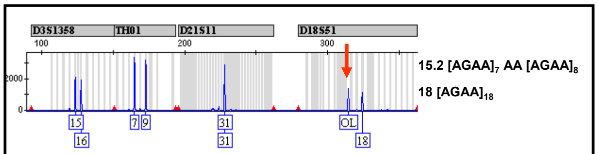


Understanding Off-Ladder Alleles

Margaret Kline
NIST Applied Genetics Group

8th Annual Advanced DNA Technical Workshop - East,
Amelia Island, FL May 20, 2009

Disclaimers

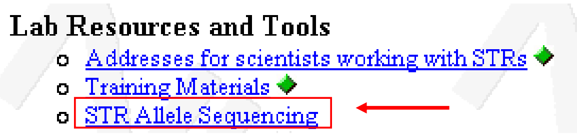
Funding: Interagency Agreement 2008-DN-R-121 between the National Institute of Justice and NIST Office of Law Enforcement Standards.

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.

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

DNA Extraction

Amplification with primers external to kit primers

Gel Cutouts with Heterozygotes

Re-Amplification

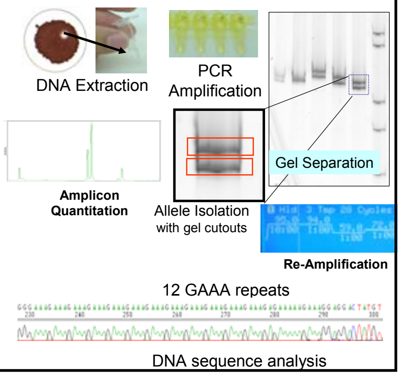
Amplicon Quantitation

ExoSAP

Cycle Sequencing

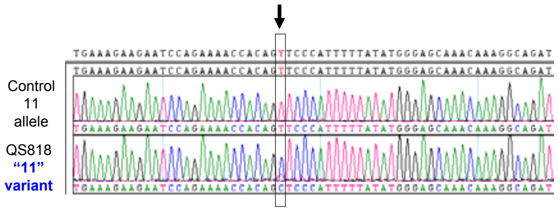
Dye Terminator Removal

F/R Sequence Alignment to Reference Sequence



DNA sequence analysis

D16S539 Variant Causing Problems during a Proficiency Test



An "11" variant allele that flip-flops between an 11 and "off-ladder" allele depending on electrophoresis conditions
(a sample in Collaborative Testing Services Test No. 07-572: Forensic Biology)

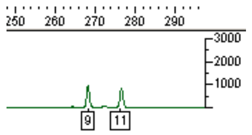
Sample has repeat region with $[GATA]_{11}$ but a base change of **T→C** 83 nucleotides upstream from the repeat (sequence shown above).

Second allele is a nominal 13 allele, $[GATA]_{13}$

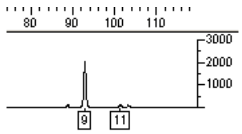
SRM 2391b Genomic 8 with D16S539

All allele calls with MiniFiler for CSF1PO, D7S820, D13S317, D18S51, D21S11, FGA, and D16S539 (with the exception noted below) ***match previously certified values.***

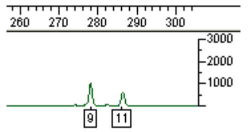
Identifiler



MiniFiler



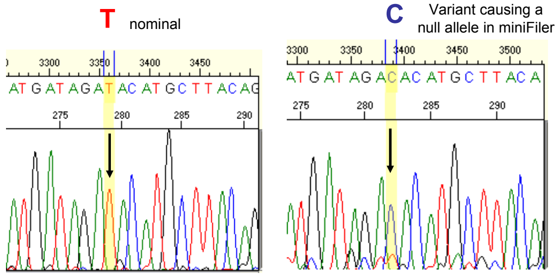
PowerPlex 16



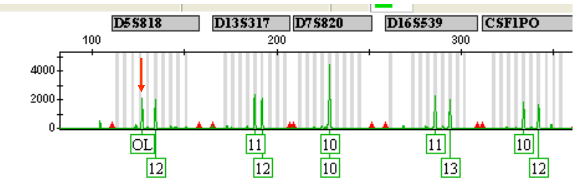
Due to primer binding site mutation

D16S539 SRM 2391b Genomic 8

This **T→C** mutation 34 bp down stream of the repeat region of D16S539 causes problems.
There an imbalance of the alleles when using PP16
Null allele when using miniFiler



D5S818 **10.1**,12



Insertion

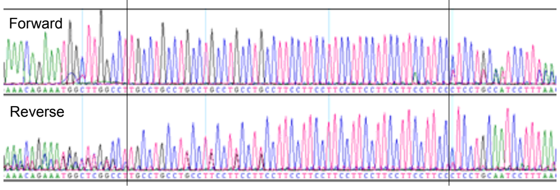
10.1 →TATTTTGA**A**$[AGAT]_{10}$

12 →TATTTTGA $[AGAT]_{12}$

10.1 allele has an **A** inserted prior to the repeat region

D2S1338 – Allele 13

$[TGCC]_6$ $[TTCC]_7$



The allelic ladder for Identifiler has alleles with 15 – 28 repeats

---

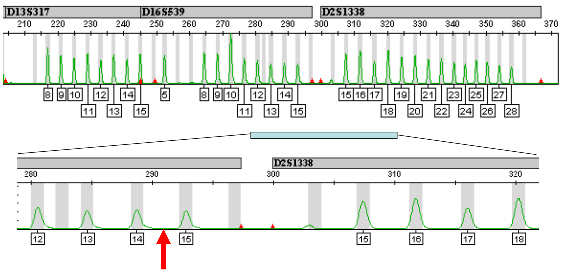
D2S1338 – 12 Allele

Variant D2S1338 Allele 12 is within the D16S539 locus

Sequencing results:

Allele 12 $[TGCC]_4$ $[TTCC]_8$

Allele 22 $[TGCC]_7$ $[TTCC]_{12}$ GTCC $[TTCC]_2$



D16S539
“14.2” = 291 bp

A state lab submitted to STRBase a new tri-allele:
D16S539 10, 12, 14.2 (Identifiler)

Likely a D2S1338 allele 11

D2S1338 alleles
11 = 291 bp
12 = 295 bp
13 = 299 bp
14 = 303 bp
15 = 307 bp

SWGDAM July 2007 (Doug Hares): search of NDIS for D16 tri-alleles with single D2 alleles found **25 profiles**

Characterizing a Variant Allele That Occurs Between Two Loci

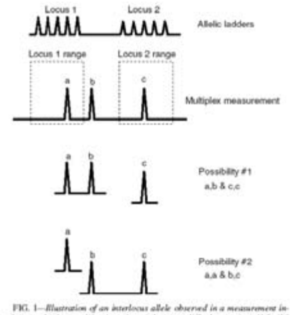
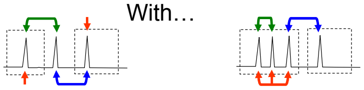


FIG. 1—*Illustration of an interlocus allele observed in a measurement involving multiplex amplification where it becomes difficult to assign allele "b" to locus 1 or locus 2.*

- Use a different multiplex STR kit with different locus combinations
- Test singleplex for each putative locus
- Example: Identifiler D16S539 and D2S1338

Butler, J.M. (2006) Genetics and genomics of core STR loci used in human identity testing. *J. Forensic Sci.* 51(2): 253-265

Steps to Detection
of Which Locus an Out-of-Range Allele Belongs With…



- Consider locus heterozygosities – heterozygote is likely from locus with higher heterozygosity (e.g., D16 = 0.766 while D2 = 0.882)
- Remember that tri-allelic patterns and homozygotes are less common than heterozygotes – thus two heterozygotes are more likely than a homozygote next to a tri-allelic pattern
- Check STRBase for variant alleles reported previously by other labs (e.g., D16 has no >16 alleles while D2 has several <15 alleles)
- Consider genotype frequencies observed for the various possible combinations (e.g., D16 11,11 = 10.7% while D2 20,20 = 0.92%)

D2S1338 – 31 Allele

Both alleles of the submitted sample were sequenced successfully.
Repeat motifs:

26 allele [TGCC]$_7$ [TTCC]$_{19}$

31 allele [TGCC]$_7$ [TTCC]$_6$ TTAC [TTCC]$_{14}$ GTCC [TTCC]$_2$

These allele repeat motifs are not currently listed on STRBase.
There is no D2S1338-31 allele currently listed on STRBase.
All other base calls for these alleles are concordant with each other.

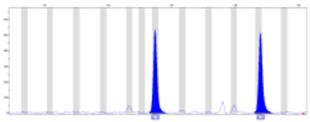
D3S1358 - 20 Allele

The 20 Allele for the D3S1358 locus is off-ladder with the AB kits (allelic ladder stops at 19). When trying to use bp sizing to call the allele it usually sizes as a 20.x.

Using PP16 amplification kit the 20 allele is included the allelic ladder so the sample bins properly. Both alleles of the submitted sample were sequenced successfully.
Repeat motifs:

16 allele TCTA TCTG [TCTA]$_{14}$

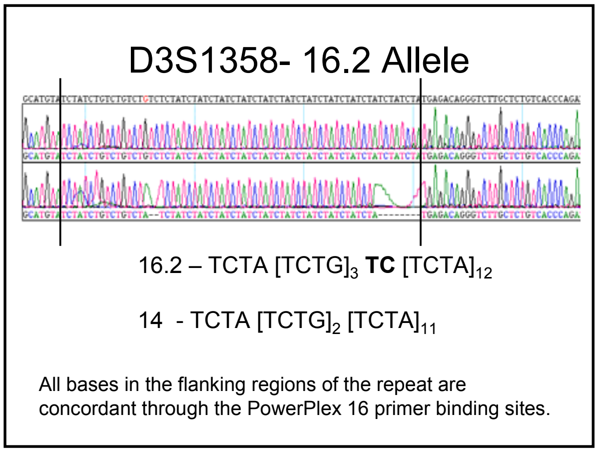
20 allele TCTA [TCTG]$_3$ [TCTA]$_{16}$



Typing the submitted sample with PP16 results in the D3S1358 type of 16,20

D3S1358 Allele size data

Allele	Average Size bp	sd	N	Difference between Alleles
11	108.6	-	1	
12	112.6	0.04	3	4.0
13	116.7	0.06	3	4.1
14	120.6	0.07	43	3.9
15	124.7	0.09	164	4.1
16	128.8	0.11	174	4.1
17	132.9	0.09	109	4.1
18	137.1	0.06	72	4.2
19	141.4	0.06	3	4.3

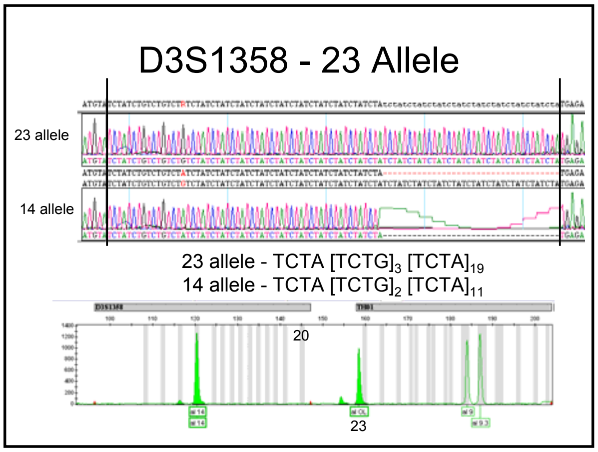


D3S1358- 16.2 Allele

16.2 – TCTA $[TCTG]_3$ **TC** $[TCTA]_{12}$

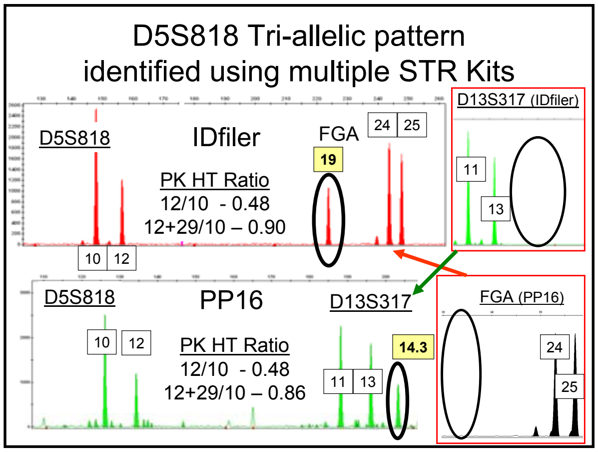
14 - TCTA $[TCTG]_2$ $[TCTA]_{11}$

All bases in the flanking regions of the repeat are concordant through the PowerPlex 16 primer binding sites.



D3S1358 - 23 Allele

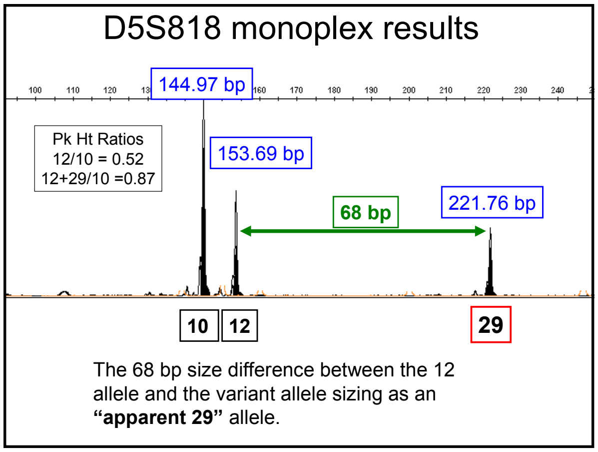
23 allele - TCTA $[TCTG]_3$ $[TCTA]_{19}$
14 allele - TCTA $[TCTG]_2$ $[TCTA]_{11}$



D5S818 Tri-allelic pattern identified using multiple STR Kits

PK HT Ratio
12/10 - 0.48
12+29/10 – 0.90

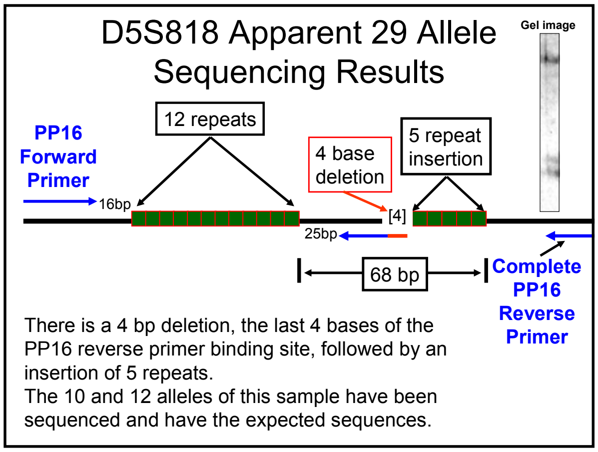
PK HT Ratio
12/10 - 0.48
12+29/10 – 0.86



D5S818 monoplex results

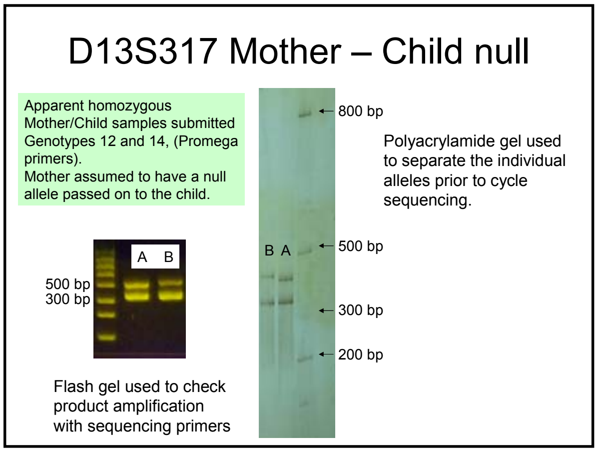
Pk Ht Ratios
12/10 = 0.52
12+29/10 =0.87

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.



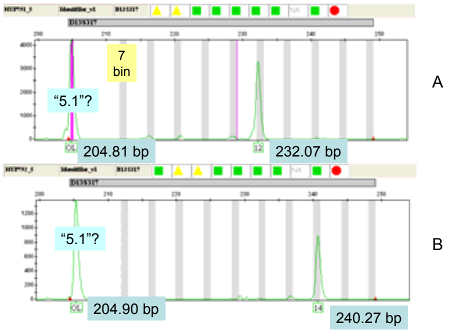
D13S317 Mother – Child null

Apparent homozygous Mother/Child samples submitted Genotypes 12 and 14, (Promega primers).
Mother assumed to have a null allele passed on to the child.

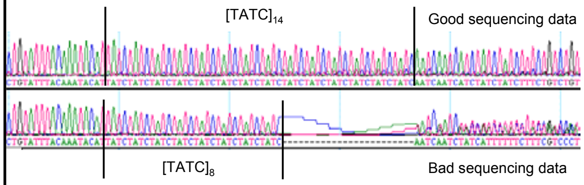
Flash gel used to check product amplification with sequencing primers

Polyacrylamide gel used to separate the individual alleles prior to cycle sequencing.

Identifiler results D13S317



Sequencing Results D13S317



Both samples, A and B, have $[TATC]_8$ repeat units in the "null" allele. Both samples have not sequenced cleanly through the area where the Promega D13S317 reverse primer anneals. Based on the AB Identiflier results for these samples there are base deletions between the end of the repeat and where the AB reverse primer anneals resulting in the apparent "5.1" allele with the AB primers.

D7S820 "8.3",10

"8.3" allele

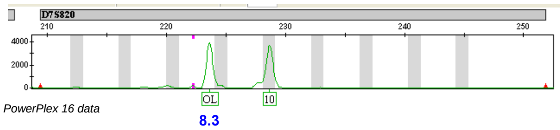
9 repeat units with a deletion 22 bases downstream from the repeat region

Deletion
A

GATA GATA GATA GATA GATA GATA GATA GATA GATA GACAGATTGATAGTTTTTTTTA_TCTC

Normal 10 allele

GATA GATA GATA GATA GATA GATA GATA GATA GATA GATA GACAGATTGATAGTTTTTTTTAATCTC

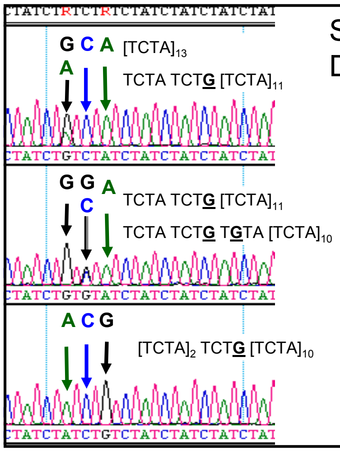


PowerPlex 16 data

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

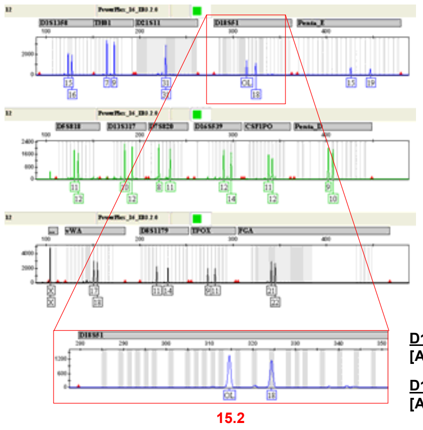
$[TCTA]_{13}$
TCTA TCT**G** $[TCTA]_{11}$
TCTA TCT**G** T**G**TA $[TCTA]_{10}$
$[TCTA]_2$ TCT**G** $[TCTA]_{10}$

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	



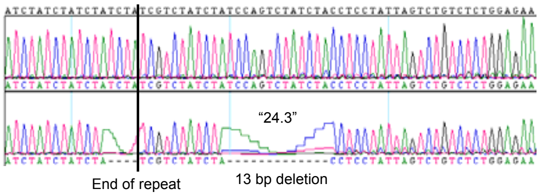
D18S51

15.2,18

D18S51 Allele 15.2
[AGAA]$_7$ AA [AGAA]$_8$

D18S51 Allele 18
[AGAA]$_{18}$

D21S11 apparent 24.3 allele



28 - [TCTA]$_4$ [TCTG]$_6$ {[TCTA]$_3$ TA [TCTA]$_3$ TCA [TCTA]$_2$ TCCATA} [TCTA]$_{10}$

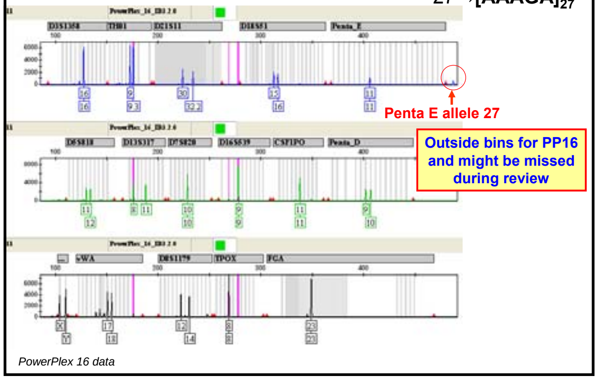
"24.3" - [TCTA]$_5$ [TCTG]$_6$ {[TCTA]$_3$ TA [TCTA]$_3$ TCA [TCTA]$_2$ TCCATA} [TCTA]$_9$ -13bp

"24.3" allele is actually a 28 allele with a 13 bp deletion 11 bases downstream the repeat region. MiniSTR primers do not bind to this allele.

Penta E 11,27

11→[AAAGA]$_{11}$

27→**[AAAGA]$_{27}$**



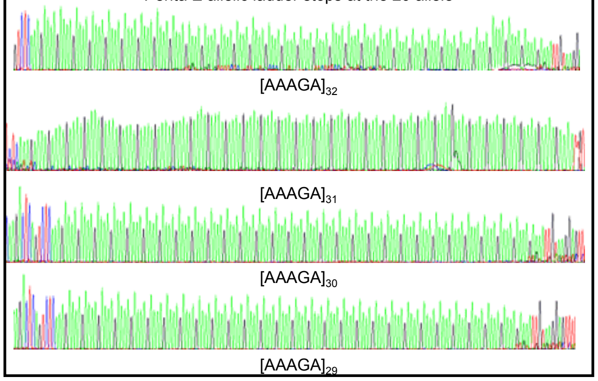
Penta E allele 27

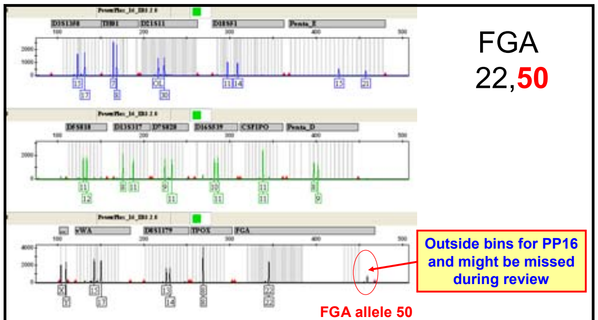
Outside bins for PP16 and might be missed during review

PowerPlex 16 data

Penta E – Large Off-ladder Alleles

Penta E allelic ladder stops at the 20 allele





FGA

22,50

Outside bins for PP16 and might be missed during review

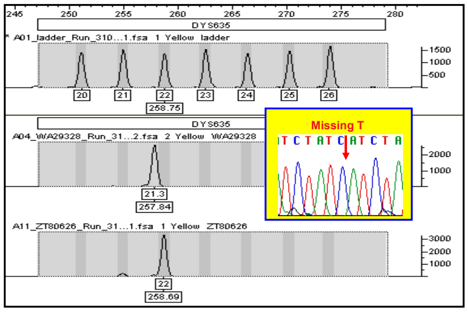
FGA allele 50

PowerPlex 16 data

FGA allele 22→[TTTC]$_3$ TTTT TTCT [CTTT]$_{14}$ CTCC [TTCC]$_2$

FGA allele 50→[TTTC]$_4$ TTTT [TTCT]$_6$ TTTT [CTTT]$_{12}$ CTGT [CTTT]$_{14}$ [CTTC]$_3$ [CTTT]$_3$ CTCC [TTCC]$_4$

An Example of an "Off-Ladder" Microvariant at the Yfiler Locus DYS635



Allele 22 bin
258.75 +/- 0.5
= 258.25 to 259.25

Allele 21.3
257.84
(-0.41 from bin)

[TCTA]$_4$(TGTA)$_2$[TCTA]$_2$(TGTA)$_2$[TCTA]$_2$(TGTA)$_2$ [TCTA]$_5$ TC-A [TCTA]$_2$

DYS389II - 29.1 Allele

Nominal 29 allele $[TCTG]_4$ $[TCTA]_{13}$ **N**$_{48}$ $[TCTG]_3$ $[TCTA]_9$

Variant 29.1 allele $[TCTG]_4$ $[TCTA]_{13}$ **N**$_{49}$ $[TCTG]_3$ $[TCTA]_9$

Or

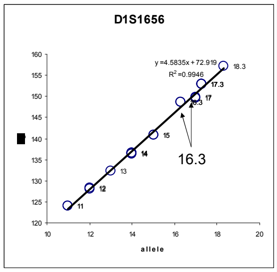
Variant 29.1 allele $[TCTG]_4$ $[TCTA]_{14}$ **N**$_{45}$ $[TCTG]_3$ $[TCTA]_9$

Variant 29.1 has an insertion of a **T** - 3 bases downstream of the repeat

New Locus D1S1656 (European)

SRM 2391b
Component Typing

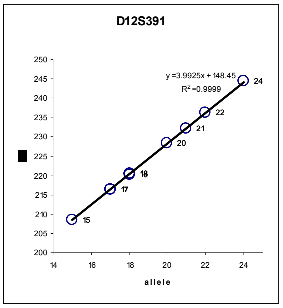
	Type	Repeat motif	Size
1	13	[TAGA]13[TG]5	132.34
	14	[TAGA]13 TAGG [TG]5	136.51
2	12	[TAGA]12[TG]5	128.15
	17.3	[TAGA]4 TGA [TAGA]12 TAGG [TG]5	153
3	14	[TAGA]13 TAGG [TG]5	136.38
	15	[TAGA]14 TAGG [TG]5	140.77
4	15	[TAGA]14 TAGG [TG]5	140.77
	17.3	[TAGA]4 TGA [TAGA]12 TAGG [TG]5	152.94
5	11	[TAGA]11[TG]5	124.12
	16.3	[TAGA]4 TGA [TAGA]11 TAGG [TG]5	148.59
6	11	[TAGA]11[TG]5	124.07
	17	[TAGA]16 TAGG [TG]5	149.64
7	12	[TAGA]12[TG]5	128.18
	17.3	[TAGA]4 TGA [TAGA]12 TAGG [TG]5	152.97
8	14	[TAGA]13 TAGG [TG]5	136.5
	16.3	[TAGA]4 TGA [TAGA]11 TAGG [TG]5	148.64
9	18.3	[TAGA]4 TGA [TAGA]13 TAGG [TG]5	157.26
10	14	[TAGA]13 TAGG [TG]5	136.5
	17	[TAGA]16 TAGG [TG]5	149.75



New Locus D12S391 (European)

SRM 2391b
Component Typing

	Type	Repeat motif	Size
1	15	[AGAT]8 [AGAC]6 AGAT	208.43
	18	[AGAT]11 [AGAC]6 AGAT	220.39
2	17	[AGAT]10 [AGAC]6 AGAT	216.36
	22	[AGAT]13 [AGAC]9	236.27
3	15	[AGAT]8 [AGAC]6 AGAT	208.36
	21	[AGAT]12 [AGAC]9	232.15
4	17	[AGAT]11 [AGAC]5 AGAT	216.25
	17	[AGAT]10 [AGAC]6 AGAT	216.25
5	18	[AGAT]11 [AGAC]6 AGAT	220.25
6	21	[AGAT]11 [AGAC]10	232.13
	22	[AGAT]12 [AGAC]10	236.19
7	17	[AGAT]10 [AGAC]6 AGAT	216.31
	20	[AGAT]10 [AGAC]9 AGAT	228.27
8	18	[AGAT]11 [AGAC]6 AGAT	220.38
	24	[AGAT]15 [AGAC]9	244.37
9	18	[AGAT]11 [AGAC]6 AGAT	220.37
	20	[AGAT]12 [AGAC]7 AGAT	228.41
10	18	[AGAT]11 [AGAC]6 AGAT	220.35
	24	[AGAT]13 [AGAC]6	244.42



Thank you for your Attention!!

Acknowledgements

Funding:

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

NIST Human Identity Project Team:

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

Email: Margaret.Kline@nist.gov