Topics and Techniques for Forensic DNA Analysis

STRs and Molecular Biology Artifacts

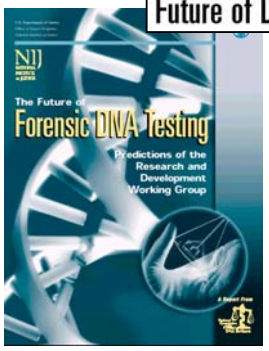
Florida Statewide Training Meeting

Indian Rocks Beach, FL
May 12-13, 2008

Dr. John M. Butler
National Institute of Standards and Technology

john.butler@nist.gov

National Commission on the Future of DNA Evidence



http://www.ojp.usdoj.gov/nij/pubs-sum/183697.htm

- Report published in Nov 2000
- Asked to estimate where DNA testing would be 2, 5, and 10 years into the future

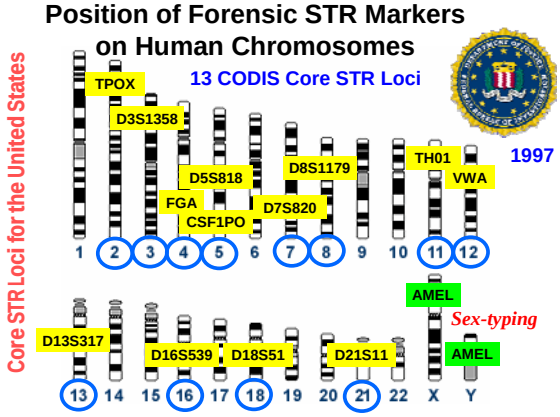
Conclusions

STR typing is here to stay for a few years because of DNA databases that have grown to contain millions of profiles

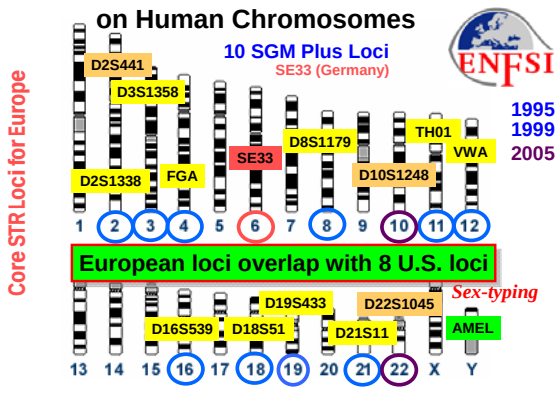
Advantages for STR Markers

- Small product sizes are generally compatible with degraded DNA and PCR enables recovery of information from small amounts of material
- Multiplex amplification with fluorescence detection enables high power of discrimination in a single test
- Commercially available in an easy to use kit format
- Uniform set of core STR loci provide capability for national and international sharing of criminal DNA profiles

Position of Forensic STR Markers on Human Chromosomes



Position of Forensic STR Markers on Human Chromosomes



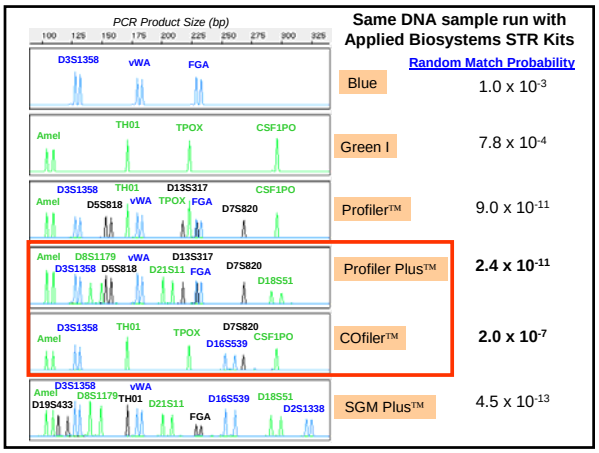
Value of STR Kits

Advantages

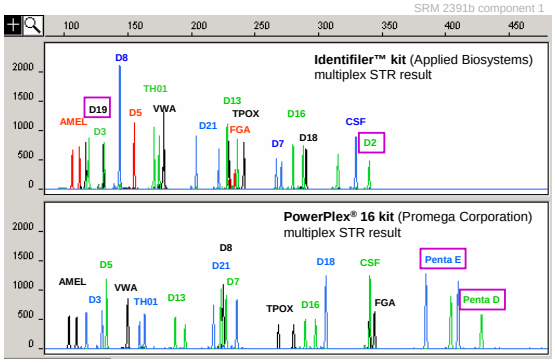
- Quality control of materials is in the hands of the manufacturer (saves time for the end-user)
- Improves consistency in results across laboratories – same allelic ladders used
- Common loci and PCR conditions used – aids DNA databasing efforts
- Simpler for the user to obtain results

Disadvantages

- Contents may not be completely known to the user (e.g., primer sequences)
- Higher cost to obtain results



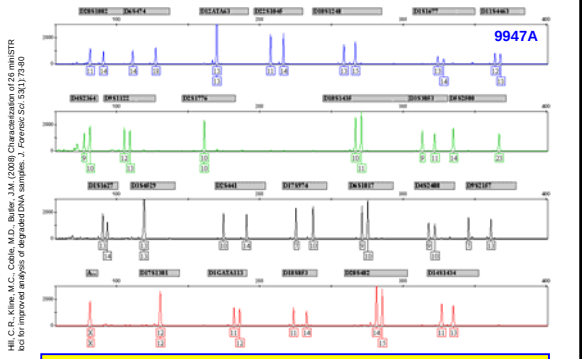
Commercial STR 16plex Kits



From Butler, J.M. (2005) Constructing STR multiplex assays. *Methods in Molecular Biology: Forensic DNA Typing Protocols* (Carracedo, A., ed.), Humana Press: Totowa, New Jersey. *297: 53-66.*

NIST "Autoplex" (26plex)

See Hill et al. AAFS 2008 talk (Washington, DC) and poster PP50 at DNA in Forensics 2008 meeting (Ancona)



Gender identification + 25 autosomal STR loci in a single amplification

How many STRs in the human genome?

- The efforts of the Human Genome Project have increased knowledge regarding the human genome, and hence there are many more STR loci available now than there were 10 years ago when the 13 CODIS core loci were selected.
- **More than 20,000 tetranucleotide STR loci have been characterized in the human genome** (Collins et al. An exhaustive DNA micro-satellite map of the human genome using high performance computing. Genomics 2003;82:10-19)
- There may be more than a million STR loci present depending on how they are counted (Ellegren H. Microsatellites: simple sequences with complex evolution. Nature Rev Genet 2004;5:435-445).
- STR sequences account for approximately 3% of the total human genome (Lander et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921).

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

Types of STR Repeat Units

Requires size based DNA separation to resolve different alleles from one another

- **Di**nucleotide (CA)(CA)(CA)(CA)
- **Tri**nucleotide (GCC)(GCC)(GCC)
- **Tetra**nucleotide (AATG)(AATG)(AATG)
- **Penta**nucleotide (AGAAA)(AGAAA)
- **Hexa**nucleotide (AGTACA)(AGTACA)

Short tandem repeat (STR) = microsatellite = simple sequence repeat (SSR)

Categories for STR Markers

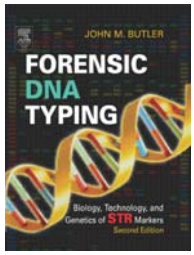
Category	Example Repeat Structure	13 CODIS Loci
Simple repeats – contain units of identical length and sequence	(GATA)(GATA)(GATA)	TPOX, CSF1PO, D5S818, D13S317, D16S539
Simple repeats with non-consensus alleles (e.g., TH01 9.3)	(GATA)(GAT-)(GATA)	TH01, D18S51, D7S820
Compound repeats – comprise two or more adjacent simple repeats	(GATA)(GATA)(GACA)	VWA, FGA, D3S1358, D8S1179
Complex repeats – contain several repeat blocks of variable unit length	(GATA)(GACA)(CA)(CATA)	D21S11

These categories were first described by **Urquhart *et al.* (1994) *Int. J. Legal Med.* 107:13-20**

Biological "Artifacts" of STR Markers

- Stutter Products
- Non-template nucleotide addition
- Microvariants
- Tri-allelic patterns
- Null alleles
- Mutations

Chapter 6 covers these topics in detail



Stutter Products

- Peaks that show up primarily one repeat less than the true allele as a result of strand slippage during DNA synthesis
- Stutter is less pronounced with larger repeat unit sizes (dinucleotides > tri- > tetra- > penta-)
- Longer repeat regions generate more stutter
- Each successive stutter product is less intense (allele > repeat-1 > repeat-2)
- Stutter peaks make mixture analysis more difficult

STR Alleles with Stutter Products

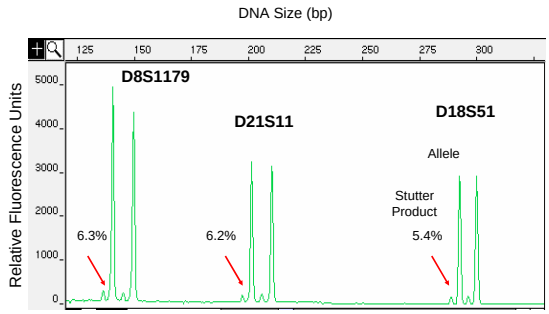
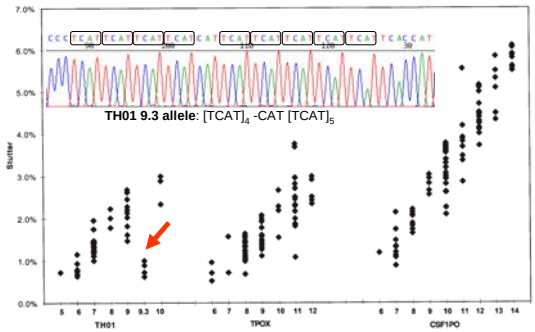


Figure 6.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Measured Stutter Percentages

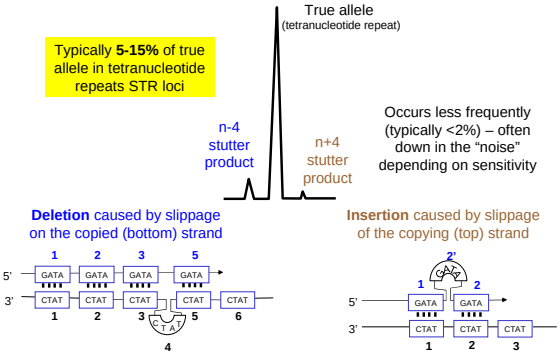
Variable by Allele Length and Composition



Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS. TWGDAM validation of AmpFISTR PCR amplification kits for forensic DNA casework. *J Forensic Sci* 2002; 47(1): 66-96.

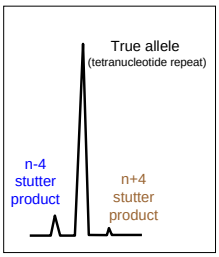
Stutter Product Formation

Repeat unit bulges out when strand breathing occurs during replication



N+4 Stutter Evaluation Summaries

- **Mass State Police DNA Lab**
- **Trying to collect data from as many laboratories as possible** to characterize N + 4 stutter percentages in various platforms.
- Please email information to **rebecca.post@pol.state.ma.us**



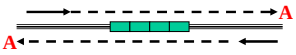
N-4 Stutter % of	main allele: allele	main allele: rfu	N+4: 'allele'	N+4: rfu	N+4 Stutter % of
6.43%	19	4664	20	57	1.22%

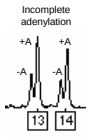
http://www.cstl.nist.gov/biotech/strbase/validation/N+4_stutter_spreadsheet.xls

Non-Template Addition

- Taq polymerase will often add an extra nucleotide to the end of a PCR product; most often an "A" (termed "adenylation")
- **Dependent on 5'-end of the reverse primer; a "G" can be put at the end of a primer to promote non-template addition**
- Can be enhanced with extension soak at the end of the PCR cycle (e.g., 15-45 min @ 60 or 72 ºC) – to give polymerase more time
- Excess amounts of DNA template in the PCR reaction can result in incomplete adenylation (not enough polymerase to go around)

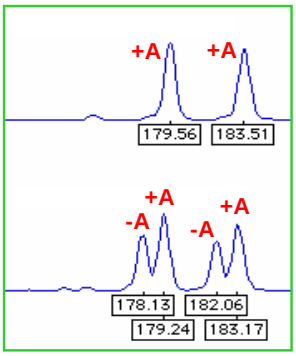
Best if there is NOT a mixture of "+/- A" peaks
(desirable to have full adenylation to avoid split peaks)





D8S1179

Impact of the 5' Nucleotide on Non-Template Addition



5'-ACAAG…

Last Base for Primer Opposite Dye Label

(PCR conditions are the same for these two samples)

5'-CCAAG…

Promega includes an ATT sequence on the 5'-end of many of their unlabeled PP16 primers to promote adenylation
see Krenke *et al.* (2002) *J. Forensic Sci.* 47(4): 773-785
http://www.cstl.nist.gov/biotech/strbase/PP16primers.htm

Higher Levels of DNA Lead to Incomplete Adenylation

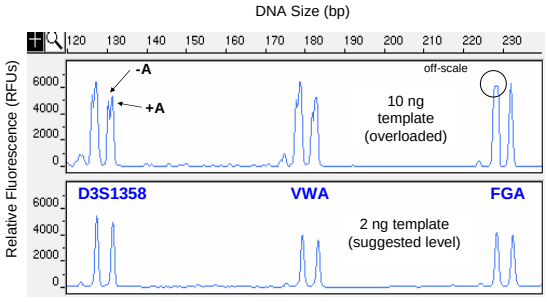
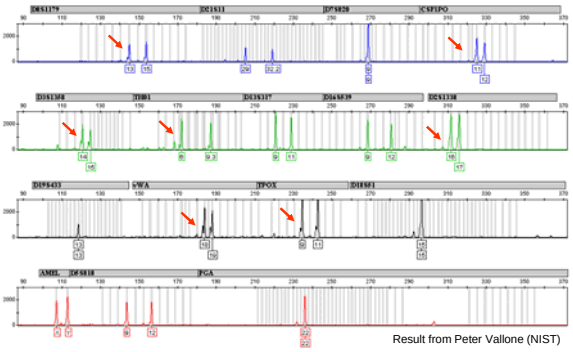


Figure 6.5, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Identifiler – Rapid PCR (36 min total time)

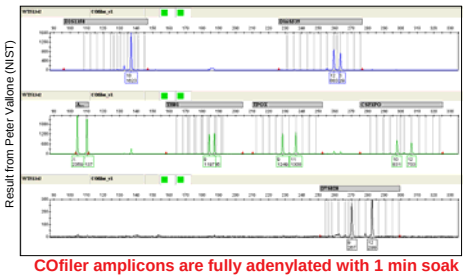
with 1 min 60 ºC adenylation soak (using different polymerases)



Result from Peter Vallone (NIST)

Rapid PCR Work and Adenylation

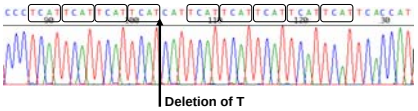
- **Poor adenylation (presence of –A peaks) is locus-specific and impacted by number of loci amplified**



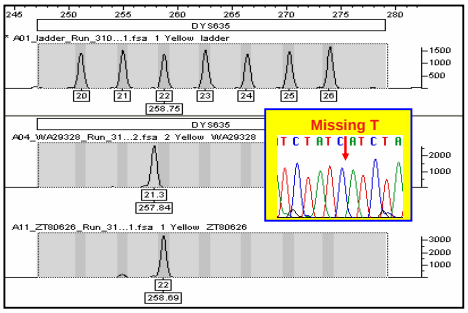
COfiler amplicons are fully adenylated with 1 min soak

Microvariant "Off-Ladder" Alleles

- Defined as alleles that are not exact multiples of the basic repeat motif or sequence variants of the repeat motif or both
- Alleles with partial repeat units are designated by the number of full repeats and then a decimal point followed by the number of bases in the partial repeat (Bar *et al. Int. J. Legal Med.* 1994, 107:159-160)
- Example: **TH01 9.3 allele**: $[TCAT]_4$ -CAT $[TCAT]_5$



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

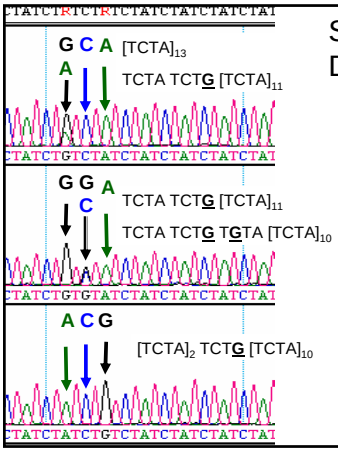


Allele 22 bin
258.75 +/- 0.5 = 258.25 to 259.25

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 $[TCTA]_2$

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.

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.

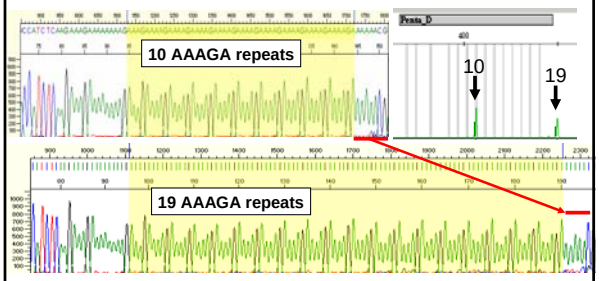
We require a minimum of 10 ng for the sequencing.
We request copies of the electropherograms demonstrating the variant allele.
The more information we have up front the better.
Please have patience we will get to your samples!

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)!

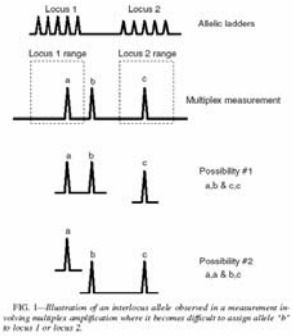
- Contact Margaret Kline: **margaret.kline@nist.gov**

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.

Characterizing a Variant Allele That Occurs Between Two Loci

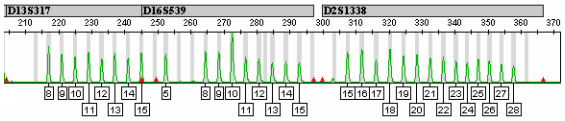


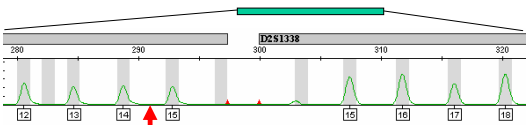
- Use a different multiplex STR kit with different locus combinations
- Test singleplex for each putative locus
- Example: Identifiler D16S539 and D2S1338

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%)





D16S539
"14.2" = 291 bp

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

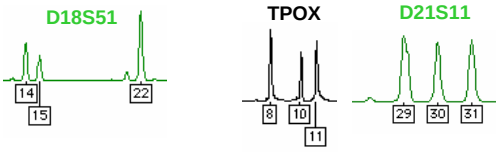
Likely a D2S1338 allele 11

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**

Three-Peak Patterns

Clayton *et al.* (2004) A genetic basis for anomalous band patterns encountered during DNA STR profiling. *J Forensic Sci.* 49(6):1207-1214



"Type 1"

Sum of heights of two of the peaks is equal to the third

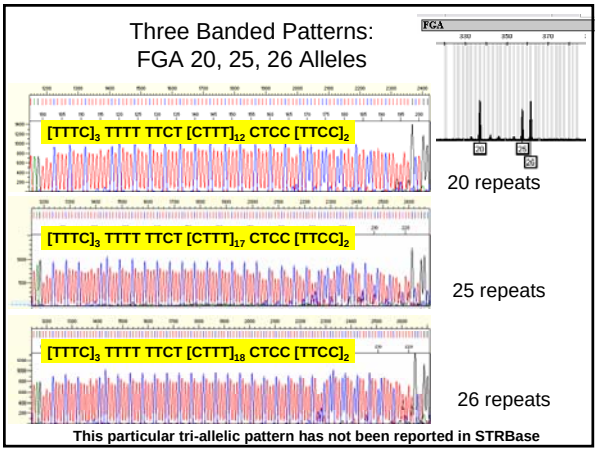
Most common in D18S51 and …..

"Type 2"

Balanced peak heights

Most common in TPOX and D21S11

Three Banded Patterns:
FGA 20, 25, 26 Alleles



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

TPOX Tri-Allelic Patterns

FSI Genetics 2008; 2(2): 134–137

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 2 (2008) 134–137

The nature of tri-allelic TPOX genotypes in African populations

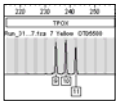
A.B. Lane

Division of Human Genetics, Room 212 James Gear Building, National Health Laboratory Service and University of the Witwatersrand, Corner of Hospital and De Korte Streets, Braamfontein, Johannesburg 2001, South Africa

Received 18 June 2007; received in revised form 8 October 2007; accepted 9 October 2007

Approximately 2.4% of indigenous South Africans have three rather than two TPOX alleles. Data collected during routine paternity testing revealed that **the extra allele is almost always allele 10** and that it segregates independently of those at the main TPOX locus. Approximately twice as many females as males have tri-allelic genotypes which suggested that **the extra allele is on an X chromosome.**

TPOX Tri-Allelic Patterns Reported on STRBase



http://www.cstl.nist.gov/biotech/strbase/var_TPOX.htm#Tri

- 6,8,**10** (4x)
- 6,9,**10** (5x)
- 6,**10**,11 (4x)
- 6,**10**,12 (1x)
- 7,8,**10** (2x)
- 7,9,**10** (1x)
- 7,**10**,11 (2x)
- 8,9,**10** (14x)
- **8,9,11 (1x)**
- 8,**10**,11 (19x)
- 8,**10**,12 (4x)
- **8,11,12 (3x)**
- 9,**10**,11 (11x)
- 9,**10**,12 (2x)
- **10**,**10**,11 (1x)
- **10**,11,12 (4x)

TPOX 10 freq
In NIST U.S. pop
Af Am 8.9%
Cau 5.6%
Hisp 3.2%

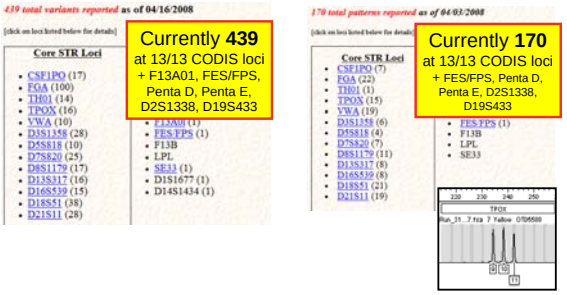
In 78 observations of 16 different TPOX tri-allelic patterns, only 4 times (5%) is allele "10" not present

Variant Alleles Cataloged in STRBase

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

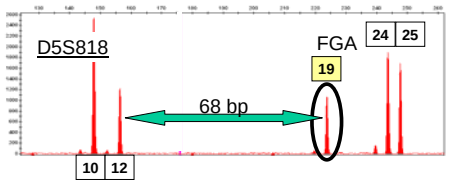
Off-Ladder Alleles

Tri-Allelic Patterns



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

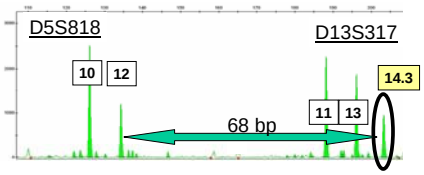
PK HT Ratio
12/10 - 0.48



PK HT Ratio
19/24 - 0.55
25/24 – 0.89

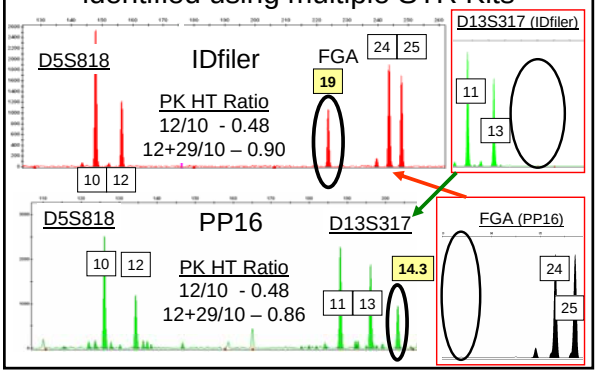
Or is this a D13S317 - Tri-allelic pattern identified using Powerplex 16?

PK HT Ratio
12/10 - 0.48

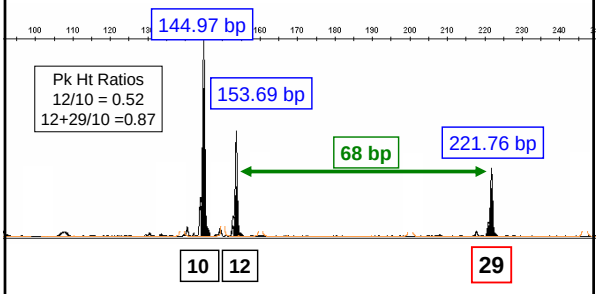


PK HT Ratio
13/11 - 0.83
14.3/11 – 0.42

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

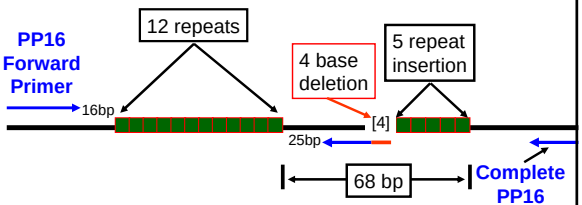


D5S818 monoplex results



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?

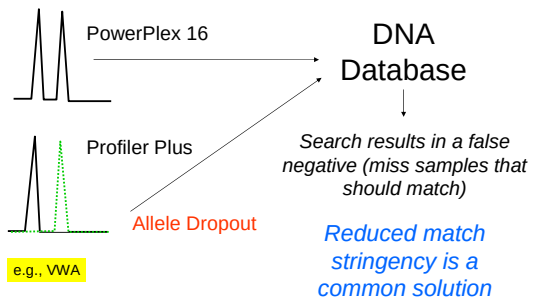
- STRBase Tri-allelic reports for FGA for 19,*,* patterns with AB amplification kits.
  - 5 reports :
  - 19,20,21; 19,20,23; 19,20,24; 19,22,23; 19,24,25
  - But there we have sequenced true tri-allelic FGA samples
- STRBase Tri-allelic reports for D13S317 for *,*, OL patterns with PP16 amplification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

Null Alleles

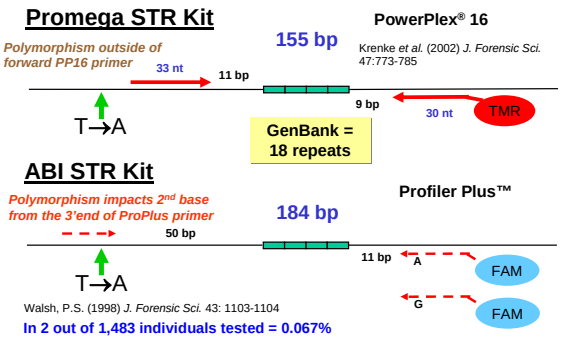
- Allele is present in the DNA sample but fails to be amplified due to a **nucleotide change in a primer binding site**
- Allele dropout is a problem because a heterozygous sample appears falsely as a homozygote
- Two PCR primer sets can yield different results on samples originating from the same source
- This phenomenon impacts DNA databases
- Large concordance studies are typically performed prior to use of new STR kits

For more information, see J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*, pp. 133-138

Concordance between STR primer sets is important for DNA databases



vWA Primer Position Comparisons



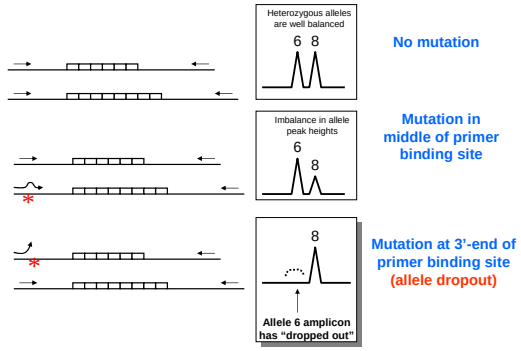
Krenke *et al.* (2002) *J. Forensic Sci.* 47:773-785

Walsh, P.S. (1998) *J. Forensic Sci.* 43: 1103-1104

In 2 out of 1,483 individuals tested = 0.067%

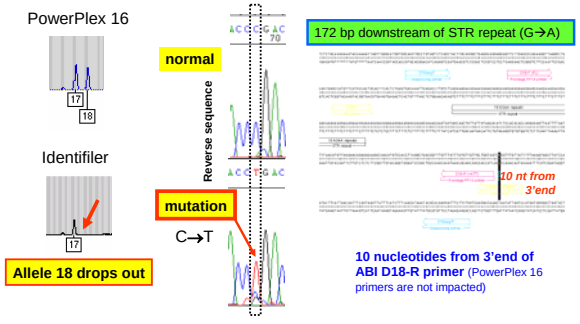
Lazaruk *et al.* (2001) *Forensic Sci Int.* 119:1-10

Impact of DNA Sequence Variation in the PCR Primer Binding Site



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 6.9, ©Elsevier Academic Press

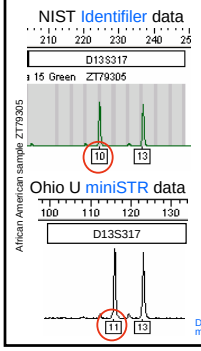
D18S51 Null Allele from Kuwait Samples with ABI Primers



Clayton et al. (2004) **Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR SGM Plus kit.** *Forensic Sci Int.* 139(2-3): 255-259

D13S317 Flanking Region Deletion

A 4 bp deletion *outside* the miniSTR primers causes the commercial kit produced allele to appear one repeat smaller…



Sequence analysis identified two regions where 4 bp deletions occur to cause this 1 repeat variation

Drabek, J., Chung, D.T., Butler, J.M., McCord, B.R. (2004) Concordance study between miniplex STR assays and a commercial STR typing kit, *J. Forensic Sci.* 49(4): 859-860.

Apparent Null Alleles Observed During Concordance Studies

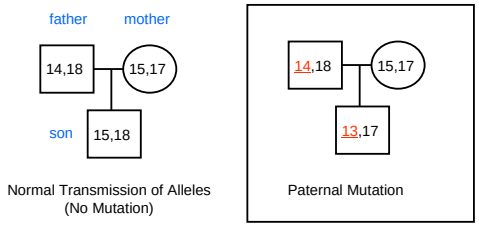
10/13 CODIS loci affected so far

New Section of STRBase (launched to track MiniFiler discordance and allele dropout frequency):
http://www.cstl.nist.gov/biotech/strbase/NullAlleles.htm

Locus	STR Kits/Assays	Results	Reference
D13S317	Identifiler vs miniplexes	Shift of alleles 10 and 11 due to deletion outside of **miniplex assay**	Butler et al. (2003), Drabek et al. (2004)
D16S539	PP1.1 vs PP16 vs COfiler	Loss of alleles with **PP1.1**; fine with PP16 and COfiler	Nelson et al. (2002)
D8S1179	PP16 vs ProPlus	Loss of alleles 15, 16, 17, and 18 with **ProPlus**; fine with PP16	Budowle et al. (2001)
FGA	PP16 vs ProPlus	Loss of allele 22 with **ProPlus**; fine with PP16	Budowle and Sprecher (2001)
D18S51	SGM vs SGM Plus	Loss of alleles 17, 18, 19, and 20 with **SGM Plus**; fine with SGM	Clayton et al. (2004)
CSF1PO	PP16 vs COfiler	Loss of allele 14 with **COfiler**; fine with PP16	Budowle et al. (2001)
TH01	PP16 vs COfiler	Loss of allele 9 with **COfiler**; fine with PP16	Budowle et al. (2001)
D21S11	PP16 vs ProPlus	Loss of allele 32.2 with **PP16**; fine with ProPlus	Budowle et al. (2001)

From Table 6.2 in J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*, p. 136

Mutation Observed in Family Trio



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 6.9, ©Academic Press

STR Measured Mutation Rates

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

STR Locus	Maternal Meioses (%)	Paternal Meioses (%)	Either Parent	Total Mutations	Rate
CSF1PO	70/179,353 (0.04)	727/504,342 (0.14)	303	1,100/683,695	**0.16%**
FGA	134/238,378 (0.06)	1,481/473,924 (0.31)	495	2,110/712,302	**0.30%**
TH01	23/189,478 (0.01)	29/346,518 (0.008)	23	75/535,996	**0.01%**
TPOX	16/299,186 (0.005)	43/328,067 (0.01)	24	83/627,253	**0.01%**
VWA	133/400,560 (0.03)	907/646,851 (0.14)	628	1,668/1,047,411	**0.16%**
D3S1358	37/244,484 (0.02)	429/336,208 (0.13)	266	732/580,692	**0.13%**
D5S818	84/316,102 (0.03)	537/468,366 (0.11)	303	924/784,468	**0.12%**
D7S820	43/334,886 (0.01)	550/461,457 (0.12)	218	811/796,343	**0.10%**
D8S1179	54/237,235 (0.02)	396/264,350 (0.15)	225	675/501,585	**0.13%**
D13S317	142/348,395 (0.04)	608/435,530 (0.14)	402	1,152/783,925	**0.15%**
D16S539	77/300,742 (0.03)	350/317,146 (0.11)	256	683/617,888	**0.11%**
D18S51	83/130,206 (0.06)	623/278,098 (0.22)	330	1,036/408,304	**0.25%**
D21S11	284/258,795 (0.11)	454/306,198 (0.15)	423	1,161/564,993	**0.21%**
Penta D	12/18,701 (0.06)	10/15,088 (0.07)	21	43/33,789	**0.13%**
Penta E	22/39,121 (0.06)	58/44,152 (0.13)	55	135/83,273	**0.16%**
D2S1338	2/25,271 (0.008)	61/81,960 (0.07)	31	94/107,231	**0.09%**
D19S433	22/28,027 (0.08)	16/38,983 (0.04)	37	75/67,010	**0.11%**
F13A01	1/10,474 (0.01)	37/65,347 (0.06)	3	41/75,821	**0.05%**
FES/FPS	3/18,918 (0.02)	79/149,028 (0.05)	None reported	82/167,946	**0.05%**
F13B	2/13,157 (0.02)	8/27,183 (0.03)	1	11/40,340	**0.03%**
LPL	0/8,821 (<0.01)	9/16,943 (0.05)	4	13/25,764	**0.05%**
SE33 (ACTBP2)	0/330 (<0.30)	330/51,610 (0.64)	None reported	330/51,940	**0.64%**

13 CODIS core loci

*Data used with permission from American Association of Blood Banks (AABB) 2002 Annual Report.

Summary of STR Mutations

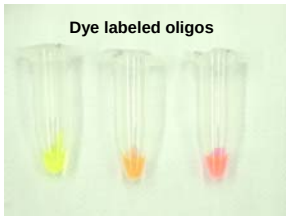
Mutations impact paternity testing and missing persons investigations but not forensic direct evidence-suspect matches…

- Mutations happen and need to be considered
- Usually 1 in ~1000 meioses
- Paternal normally higher than maternal
- VWA, FGA, and D18S51 have highest levels
- TH01, TPOX, and D16S539 have lowest levels

Primer Synthesis and Dye Blobs

- Oligonucleotide primers are synthesized from a 3'-to-5' direction on solid-phase supports using phosphoramidite chemistry
- The fluorescent dye is attached at 5'end of the primer (it is the last component added)
- The coupling reaction at each step of primer synthesis is not 100%, which can lead to some minor level impurities
- Left-over dye molecules that are not removed by post-synthesis purification can be carried through the PCR amplification step and injected onto the capillary to produce "dye blobs" or "dye artifacts" in CE electropherograms (wider than true allele peaks)

PCR Primer Quality Control

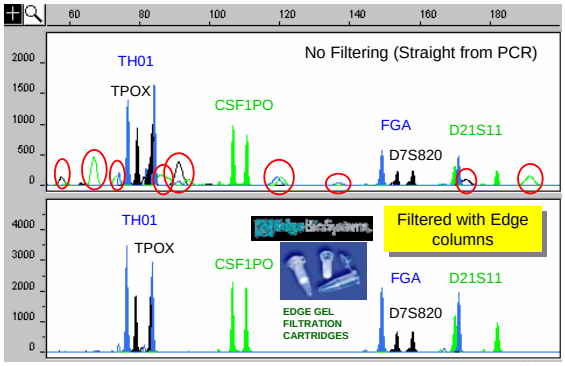


6FAM (yellow), **VIC** (orange), **NED** (red)

- UV Spec to determine concentration
- HPLC to evaluate purity
- TOF-MS to confirm correct sequence
- CE (ABI 310) to determine presence of residual dye molecules ("dye blobs")

Butler *et al.* (2001) *Forensic Sci. Int.* 119: 87-96

Problems with Dye Artifacts from Fluorescent Primers



STRBase

Short Tandem Repeat DNA Internet Database

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

General Information
- Intro to STRs (downloadable PowerPoint)
- STR Fact Sheets
- Sequence Information
- Multiplex STR Kits
- Variant Allele Reports
- Training Slides

Forensic Interest Data
- FBI CODIS Core Loci
- DAB Standards
- NIST SRMs 2391
- Published PCR Primers
- Y-Chromosome STRs
- Population Data
- Validation Studies
- miniSTRs

Supplemental Info
- Reference List **>3000**
- Technology Review
- Addresses for Scientists
- Links to Other Web Sites
- DNA Quantitation
- mtDNA
- New STRs

New information is added regularly…

Thank you for your attention…

Funding from the **National Institute of Justice (NIJ)** through NIST Office of Law Enforcement Standards


Margaret Kline
STR allele sequencing


Jan Redman
Variant allele cataloging


Becky Hill
miniSTRs and 26plex work


Amy Decker
Y-STRs


Pete Vallone
Rapid PCR

http://www.cstl.nist.gov/biotech/strbase
john.butler@nist.gov
301-975-4049

miniSTR Collaborators
Bruce McCord (FIU)
Mike Coble (AFDIL)