The Copenhagen Forensic Genetic Summer School
Advanced Topics in STR DNA Analysis
June 27-28, 2012

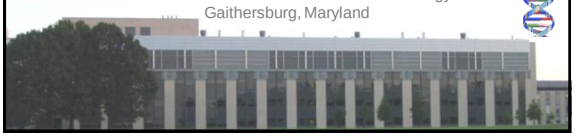
University of Copenhagen

NIST National Institute of Standards and Technology

STR Biology and Artifacts

Mike Coble and Becky Hill

NIST Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland



Limitations of DNA Testing (Past)

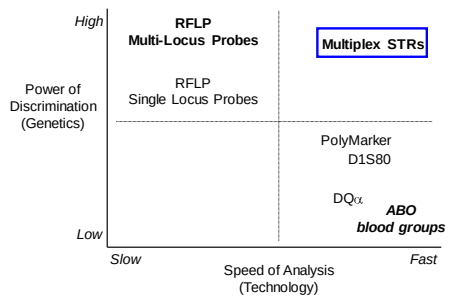


Figure 1.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

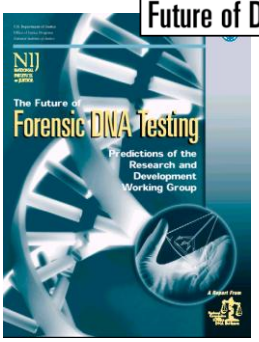
STRBase

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

Outline

- Timeline of field and growth of STR use
- STR characteristics and biology
- STR core loci and commonly used kits
- Biology of STRs: Stutter, microvariants, null alleles
- Multiplex Kits

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

ENFSI Compiled Data

Country	Population size	Persons				Stains	Matches						
		A	S	CO	T		Person/Stain			Stain/Stain	Total	Date
							S	CO	T			
Austria	8.100.000				145.527	44.831			13.902	6.084	19.986	jun-11
Belgium	10.400.000		n/a	21.886	21.886	24.624	291	1.417	1.708	2.872	4.580	jun-11
Bulgaria	7.900.000				17.618	1.147			377	122	499	jul-09
Croatia	4.600.000				29.293	5.180			3.752	1.519	5.271	feb-10
Cyprus	772.000											
Czech Republic	10.300.000		1.681	73.496	75.177	14.061			5.855	2.638	8.493	jun-11
Denmark	5.500.000				73.908	40.275			19.579	-	19.579	sep-11
Estonia	1.400.000				29.274	9.376			2.860	895	3.755	jul-09
Finland	5.385.000				111.991	12.422			14.032	1.566	15.598	jun-11
France	59.300.000	118.637	1.345.274	352.853	1.698.127	103.792	34.496	11.198	45.694	6.734	52.428	jul-11
Georgia	4.700.000											
Germany	81.835.000				711.159	184.782			86.422	23.841	110.263	jun-11
Switzerland	7.779.000				126.663	33.109			28.997	7.034	36.031	jun-11
Turkey	66.800.000											
UK (England & Wales)	53.700.000				5.368.950	387.563			1.422.573	235.892	1.658.465	jun-11
Ukraine	47.600.000											
Total	**770.487.000**				**9.300.393**	**1.025.895**			**1.745.774**	**337.544**	**2.085.902**	

European Network of Forensic Science Institutes

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

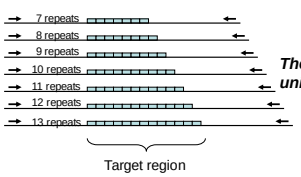
STR Biology

Short Tandem Repeat (STR) Markers

An accordion-like DNA sequence that occurs between genes

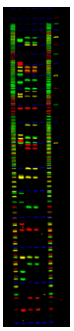
TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGAAGACA
GGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**
GATAGATATCATTGAAAGACAAAACAGAGATGGATGATAGATACA
TGCTTACAGATGCACAC

= 12 GATA repeats ("12" is all that is reported)

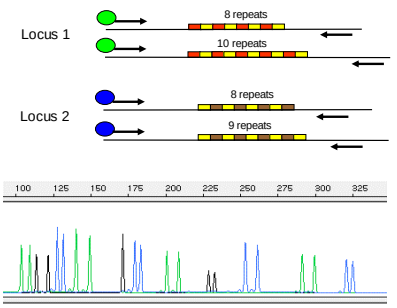


The number of consecutive repeat units can vary between people

The polymerase chain reaction (PCR) is used to amplify STR regions and label the amplicons with fluorescent dyes using locus-specific primers



Scanned Gel Image



Capillary Electropherogram

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**

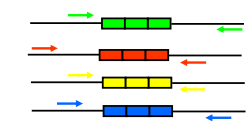
How many STRs in the human genome?

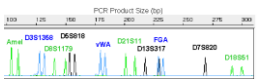
- The efforts of the Human Genome Project have increased knowledge regarding the human genome, and hence there are many more STR loci available now than there were 10 years ago when the 13 CODIS core loci were selected.
- **More than 20,000 tetranucleotide STR loci have been characterized in the human 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.

Multiplex PCR

(Parallel Sample Processing)



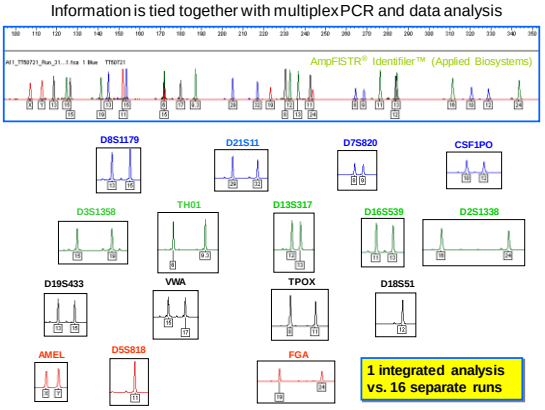


- **Compatible primers are the key to successful multiplex PCR**
- **STR kits are commercially available**
- **15 or more STR loci can be simultaneously amplified**

Challenges to Multiplexing
- primer design to find compatible primers (no program exists)
- reaction optimization is highly empirical often taking months

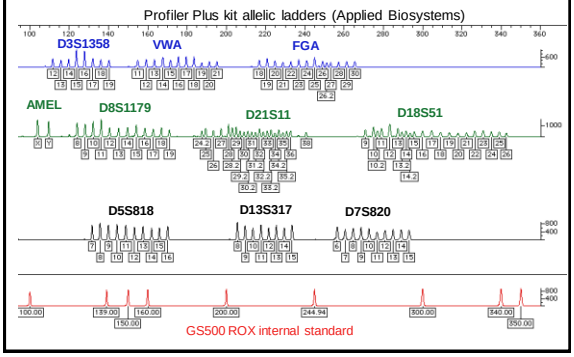
Advantages of Multiplex PCR
- –Increases information obtained per unit time (increases power of discrimination)
- –Reduces labor to obtain results
- –Reduces template required (smaller sample consumed)

Information is tied together with multiplex PCR and data analysis



1 integrated analysis vs. 16 separate runs

Companies Supply **Allelic Ladders** in STR Kits to **Aid Interlaboratory Consistency**



Biological "Artifacts" of STR Markers

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

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-)

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)

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

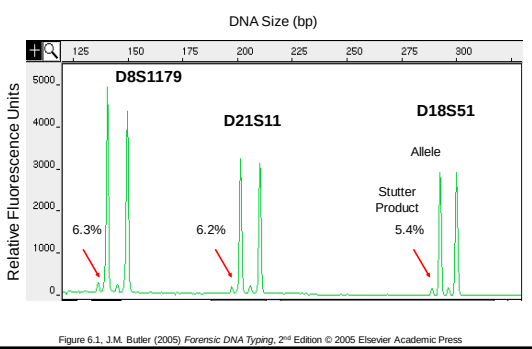
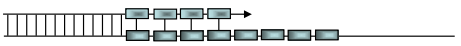


Figure 6.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Slipped Strand Mispairing Model

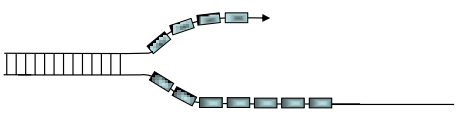
Step 1



Taq DNA Polymerase has extended through 4 repeat units

Slipped Strand Mispairing Model

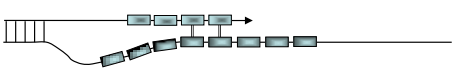
Step 2



Taq has fallen off allowing the two strands to breathe apart.

Slipped Strand Mispairing Model

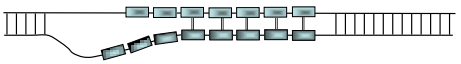
Step 3



When the two strands re-anneal the template (bottom) strand has looped out and the extending strand aligns out-of-register by one repeat unit.

Slipped Strand Mispairing Model

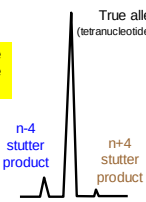
Step 4



The newly completed strand contains only 7 repeat units, while the template strand has the original 8 repeat units.

Stutter Product Formation

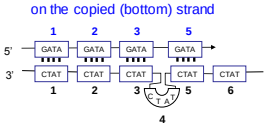
Repeat unit bulges out when strand breathing occurs during replication



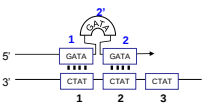
Typically **5-15%** of true allele in tetranucleotide repeats STR loci

Occurs less frequently (typically <2%) – often down in the "noise" depending on sensitivity

Deletion caused by slippage on the copied (bottom) strand



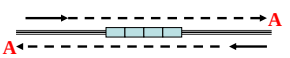
Insertion caused by slippage of the copying (top) strand

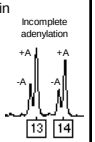


Non-Template Addition

- 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)

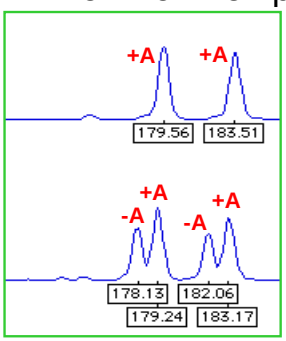
Best if there is NOT a mixture of "+/- A" peaks
(desirable to have full adenylation to avoid split peaks)





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

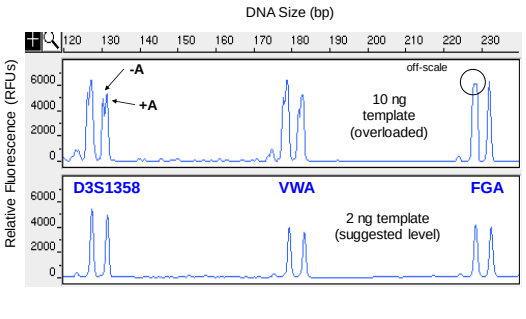


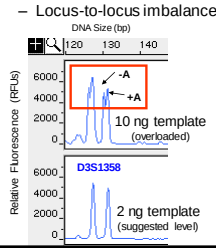
Figure 6.5, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Impact of DNA Amount into PCR

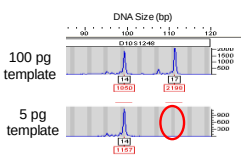
Reason that DNA Quantitation is Important Prior to Multiplex Amplification

Generally 0.5 – 2.0 ng DNA template is best for STR kits

- Too much DNA
  - Off-scale peaks
  - Split peaks (+/-A)
  - Locus-to-locus imbalance



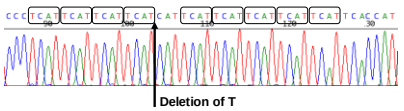
- Too little DNA
  - Heterozygote peak imbalance
  - Allele drop-out
  - Locus-to-locus imbalance



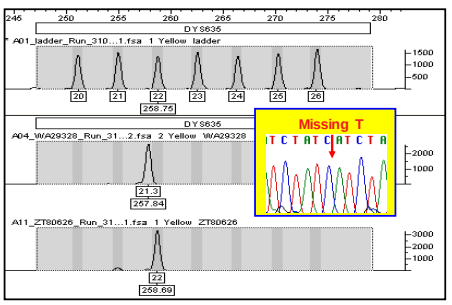
Stochastic effect when amplifying low levels of DNA produces allele dropout

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



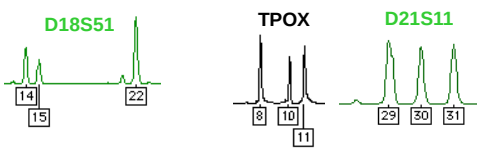
Allele 22 bin
258.75 +/- 0.5 = 258.25 to 259.25

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 $[TCTA]_2$

Tri-Allele 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

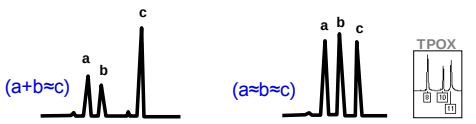
"Type 2"

Balanced peak heights

Most common in TPOX and D21S11

Tri-Allelic Patterns

- Tri-alleles are Copy Number Variants (CNVs) in the human genome detected as three peaks at a single locus rather than the expected single (homozygous) or double (heterozygous) peak



Type 1

Sum of peak heights for two alleles is almost equal to the third allele

Type 2

Fairly balanced peak heights are observed

Slide from Steven Myers, CA DOJ Data from Missouri Highway Patrol DNA Lab

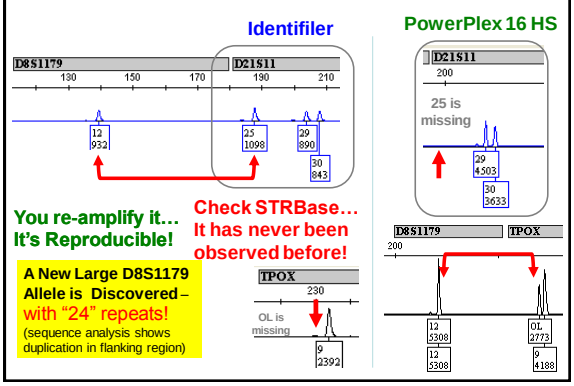
Frequency of Tri-Allelic Patterns

- Database Size: **69,000**
- Overall Average Occurrence: **1 in 1,000**

Note: This is Steven's summary of Missouri's data. This table is not on STRBase.

Locus	Observations	1 in...
D3S1358	2	35,000
VWA	10	6,900
FGA	11	6,300
D8S1179	2	35,000
D21S11	9	7,700
D18S51	3	23,000
D5S818	1	69,000
D13S317	4	17,000
D7S820	0	
D16S539	3	23,000
TH01	0	
TPOX	9	7,700
CSF1PO	1	69,000
Penta D	3	23,000
Penta E	10	6,900
Combined	68	1,000

How Do You Characterize Your Tri-Allelic Patterns?



You re-amplify it… It's Reproducible!

Check STRBase… It has never been observed before!

A New Large D8S1179 Allele is Discovered – with "24" repeats! (sequence analysis shows duplication in flanking region)

D8S1179

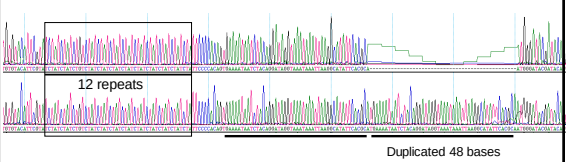
All Previously Known Alleles

Many alleles sequences are not known

We just set the new world record for the largest D8 allele (24)

Allele (Repeat #)	Promega PowerPlex 16	ABI Identifiler	Repeat Structure [TCTR]$_n$	Reference
6	199 bp	119 bp	Not published	STRBase
7	203 bp	123 bp	$[TCTA]_7$	Griffiths *et al.* (1998)
8	207 bp	127 bp	$[TCTA]_8$	Barber and Parkin (1996)
9	211 bp	131 bp	$[TCTA]_9$	Barber and Parkin (1996)
10	215 bp	135 bp	$[TCTA]_{10}$	Barber and Parkin (1996)
10.1	216 bp	136 bp	Not published	STRBase
10.2	217 bp	137 bp	Not published	STRBase
11	219 bp	139 bp	$[TCTA]_{11}$	Barber and Parkin (1996)
12	223 bp	143 bp	$[TCTA]_{12}$	Barber and Parkin (1996)
12.1	224 bp	144 bp	Not published	STRBase
12.2	225 bp	145 bp	Not published	STRBase
12.3	226 bp	146 bp	Not published	STRBase
13 (a)	227 bp	147 bp	$[TCTA]_4[TCTG]_1[TCTA]_{11}$	Barber and Parkin (1996)
13 (b)	227 bp	147 bp	$[TCTA]_3[TCTG]_1[TCTA]_{13}$	**Kline *et al.* (2010)**
13 (c)	227 bp	147 bp	$[TCTA]_4[TCTG]_1TGTA[TCTA]_{10}$	**Kline *et al.* (2010)**
13 (d)	227 bp	147 bp	$[TCTA]_{13}$	**Kline *et al.* (2010)**
13.1	228 bp	148 bp	Not published	STRBase
13.2	229 bp	149 bp	Not published	STRBase
13.3	230 bp	150 bp	Not published	STRBase
14	231 bp	151 bp	$[TCTA]_3[TCTG]_1[TCTA]_{11}$	Barber and Parkin (1996)
14.1	232 bp	152 bp	Not published	STRBase
14.2	233 bp	153 bp	Not published	STRBase
15	235 bp	155 bp	$[TCTA]_3[TCTG]_1[TCTA]_{12}$	Barber and Parkin (1996)
15.1	236 bp	156 bp	Not published	STRBase
15.2	237 bp	157 bp	Not published	STRBase
15.3	238 bp	158 bp	Not published	STRBase
16	239 bp	159 bp	$[TCTA]_3[TCTG]_1[TCTA]_{13}$	Barber and Parkin (1996)
16.1	240 bp	160 bp	Not published	STRBase
17	243 bp	163 bp	$[TCTA]_4[TCTG]_1[TCTA]_{13}$	Barber and Parkin (1996)
17.1	244 bp	164 bp	Not published	STRBase
17.2	245 bp	165 bp	Not published	STRBase
18	247 bp	167 bp	$[TCTA]_3[TCTG]_1[TCTA]_{15}$	Barber and Parkin (1996)
19	251 bp	171 bp	$[TCTA]_4[TCTG]_1[TCTA]_{15}$	Griffiths *et al.* (1998)
20	255 bp	175 bp	Not published	**STRBase**

D8S1179 12,"24"



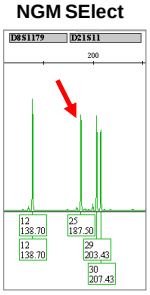
Allele 12 : $[TCTA]_2$ TCTG $[TCTA]_9$

Allele "24" : $[TCTA]_2$ TCTG $[TCTA]_9$ **duplication of the 48 bases 10 bases downstream of the repeat**

Result with This Large D8S1179 Allele

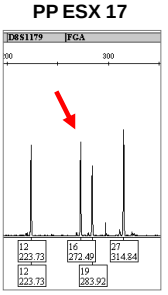
Using European STR Kits

NGM SElect



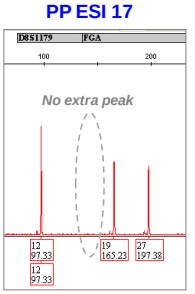
False D21S11 tri-allele

PP ESX 17



False FGA tri-allele

PP ESI 17

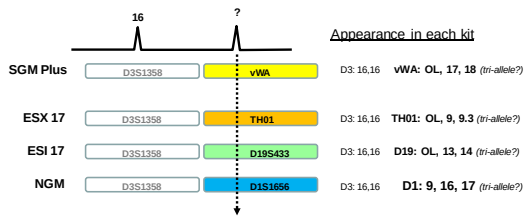


Reverse primer internal to duplicated flanking region

Recommendations for Tri-Allelic Patterns

- **Re-injecting a sample with the same STR kit does not help answer the question**
- **Run a different STR kit with loci in different configurations**
- This duplicate testing will help confirm that you have a true tri-allele rather than an extremely small or large allele that is out of the STR kit defined allele bins for a locus
- **Recording tri-allelic patterns correctly improves database searching comparability when states are using different STR kits**

Tale of a Large D3S1358 Allele



22 repeats from sequence analysis
$TCTA(TCTG)_3(TCTA)_4TCTG(TCTA)_{13}$

D3S1358 Allele 22
Initially measured as 22.1 or 22.2

Raziel et al. (2012) *FSI Genetics* 6: 108-112

Variant STR Allele Sequencing


Margaret Kline

Main Points:

- **STR allele sequencing has been provided free to the community** for the past ten years thanks to NIJ-funding
- Article provides primer sequences (outside of all known kit primers) for 23 autosomal STRs & 17 Y-STRs and full protocol for gel separations and sequencing reactions
  - 111 normal and variant alleles sequenced (at 19 STR & 4 Y-STRs)
  - 17 null alleles sequenced (with impact on various STR kit primers)


Contents lists available at ScienceDirect
Forensic Science International: Genetics
journal homepage: www.elsevier.com/locate/fsig
Short communication
STR sequence analysis for characterizing normal, variant, and null alleles
Margaret C. Kline *, Carolyn R. Hill, Amy E. Decker [1], John M. Butler

Presentations/Publications:

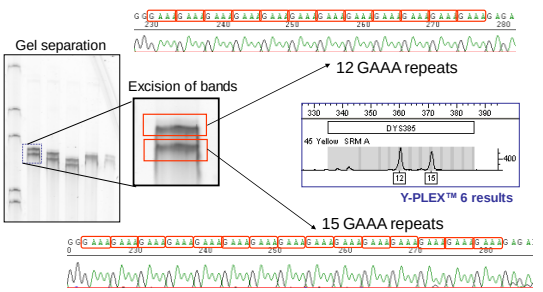
- FSI Genetics article (Aug 2011) and numerous talks

Sequencing Variant Alleles

- "Off-ladder" variants, null alleles, or any other "odd" result seen in datasets
- Sample sequencing free of charge
  - NIJ funds this work
  - 10 ng genomic DNA sample requirement with an electropherogram of the result and specified marker in question
- Results provided to customer and listed on STRBase:

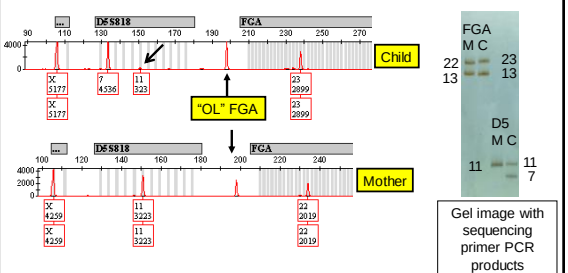
http://www.cstl.nist.gov/biotech/strbase/STRseq.htm

Sequencing Individual Heterozygous (DYS385) Alleles



Kline, M.C., Hill, C.R., Decker, A.E., Butler, J.M. (2011) STR sequence analysis for characterizing normal, variant, and null alleles. Forensic Sci. Int. Genet. 5(4): 329-332

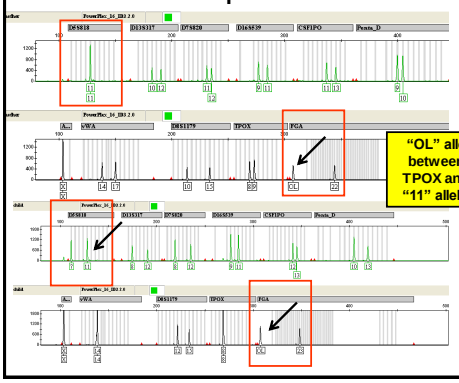
Example: Mother/Child Samples Identifiler typing



"OL" allele present for both mother and child between FGA and D5S818 with a weak "11" allele at D5S818

Jiang, W., Kline, M., Hu, P., Wang, Y. (2011) Identification of dual false indirect exclusions on the D5S818 and FGA loci. Legal Medicine 13: 30-34

Powerplex 16 results

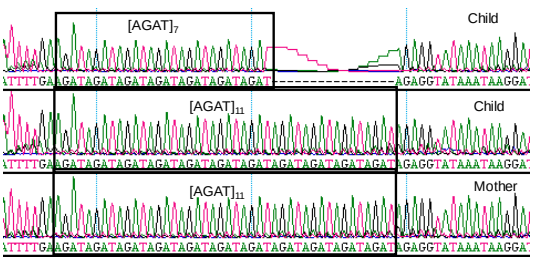


Mother
D5S818: 11,11
FGA: OL,22

"OL" allele is now between FGA and TPOX and the Child "11" allele is strong

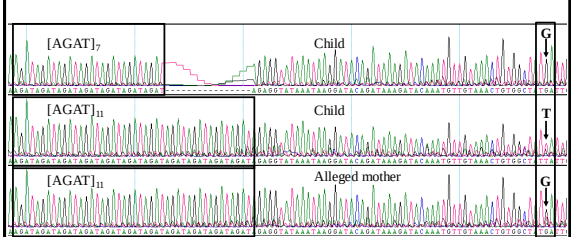
Child
D5S818: 7,11
FGA: OL,23

D5S818 Mother-Child



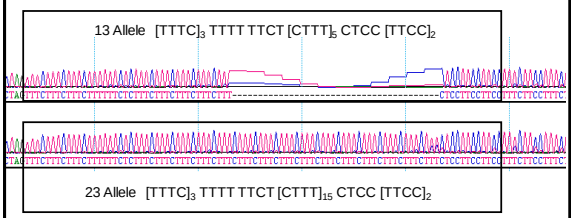
Child types as 7,11 with Powerplex 16 primers
Mother types as 11,11 with Powerplex 16 primers
Both types confirmed by sequencing

D5S818 Mother-Child



Child has a null 11 allele with Identifiler primers,
G→T SNP was found 55 bp downstream from the repeat.

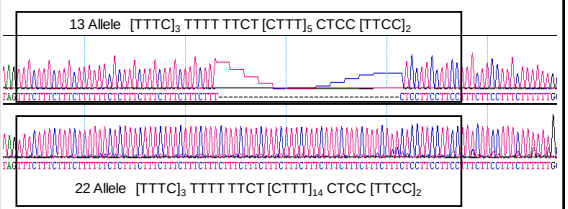
Child - FGA 13,23



Powerplex 16 Ladder for FGA ends at 16 allele

Note: When running Identifiler the 13 allele is not annotated With PP16 the 13 allele is noted as "OL".

Mother - FGA 13,22



Powerplex 16 Ladder for FGA ends at 16 allele

Tri-Alleles Cataloged in STRBase

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

Core STR Loci (220)

- CSF1PO (8)
- FGA (32)
- TH01 (4)
- TPOX (18)
- VWA (23)
- D3S1358 (10)
- D5S818 (8)
- D7S820 (11)
- D8S1179 (16)
- D13S317 (15)
- D16S539 (12)
- D18S51 (30)
- D21S11 (23)

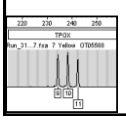
Other Common STR Loci (39)

- D2S1338 (3)
- D19S433 (7)
- Penta D (12)
- Penta E (14)
- F13A01
- FES/FPS (1)
- F13B
- LPL
- SE33
- D10S1248 (1)
- D12S391 (1)
- D2S441

Y-STR Loci (54)

duplications or triplications

- DYS19 (7)
- DYS389I (4)
- DYS389II (6)
- DYS390 (1)
- DYS391 (3)
- DYS392
- DYS393 (2)
- DYS385 a/b (18)
- DYS438 (4)
- DYS439 (5)
- DYS437 (2)
- DYS448 (1)
- DYS456 (1)
- DYS458
- DYS635/GATA-C4
- Y-GATA-H4



313 total

as of 04/12

Variant Alleles Cataloged in STRBase

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

Core STR Loci (403)

- CSF1PO (22)
- FGA (109)
- TH01 (20)
- TPOX (21)
- VWA (14)
- D3S1358 (30)
- D5S818 (17)
- D7S820 (26)
- D8S1179 (22)
- D13S317 (18)
- D16S539 (21)
- D18S51 (47)
- D21S11 (40)

Other Common STR Loci (170)

- D2S1338 (27)
- D19S433 (31)
- Penta D (38)
- Penta E (30)
- D12S391 (8)
- D1S1656 (6)
- D2S441 (4)
- D10S1248
- D22S1045
- SE33 (21)
- D6S1043
- F13A01 (2)
- FES/FPS (1)
- F13B
- LPL
- D1S1677 (1)
- D14S1434 (1)

Y-STR Loci (64)

- DYS19 (3)
- DYS389I (3)
- DYS389II (1)
- DYS390 (2)
- DYS391
- DYS392 (4)
- DYS393 (1)
- DYS385 a/b (20)
- DYS438 (3)
- DYS439 (4)
- DYS437 (3)
- DYS448 (1)
- DYS456 (4)
- DYS458 (10)
- DYS635 (3)
- Y-GATA-H4 (2)

638 total

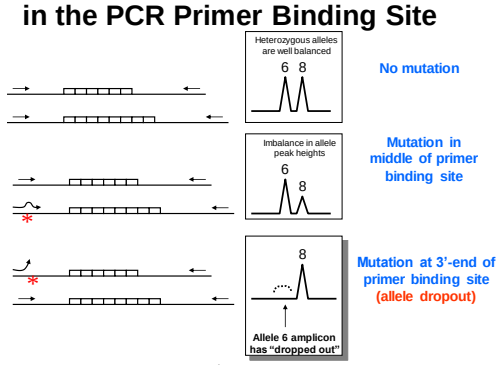
as of 04/12

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

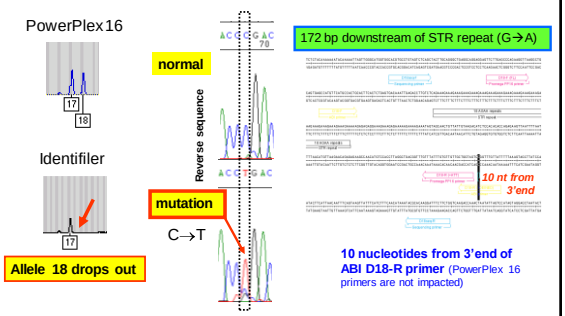
For more information, see J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*, pp. 133-138

Impact of DNA Sequence Variation in the PCR Primer Binding Site

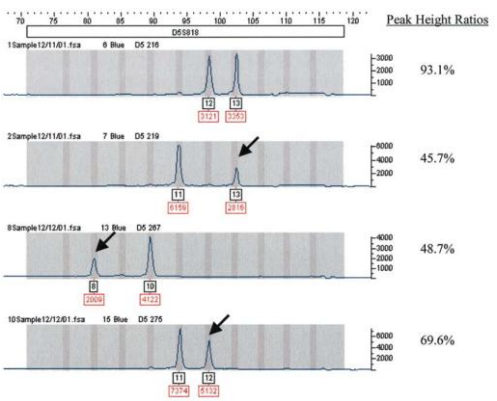


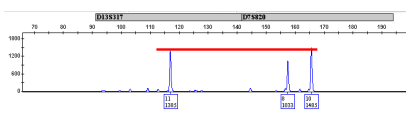
Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 6.9, ©Elsevier Academic Press

D18S51 Null Allele from Kuwait Samples with ABI Primers



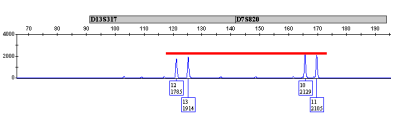
Clayton et al. (2004) **Primer binding site mutations affecting the typing of STR loci contained within the AMPFISTR SGM Plus kit.** *Forensic Sci Int.* 139(2-3): 255-259



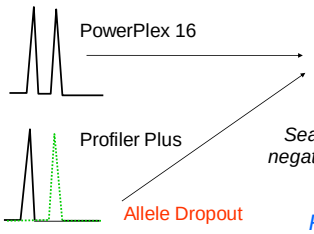


"Apparent" Homozygote – note peak height is similar to heterozygote

Null Allele



Concordance between STR primer sets is important for DNA databases

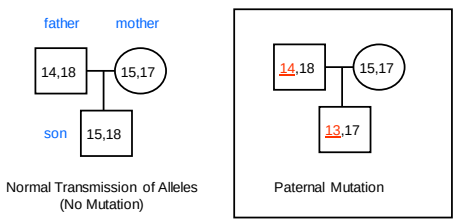


DNA Database

Search results in a false negative (miss samples that should match)

Reduced match stringency is a common solution

Mutation Observed in Family Trio



J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*

STR Measured Mutation Rates

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

Apparent Mutations Observed at STR Loci in the Course of Paternity Testing*

13 CODIS core loci

STR System	Maternal Meioses (%)	Paternal Meioses (%)	Number from either	Total Number of Mutations	Mutation Rate
CSF1PO	93/304,307 (0.03)	982/640,118 (0.15)	410	1,487/947,425	**0.16%**
FGA	205/408,230 (0.05)	2,210/692,776 (0.32)	710	3,125/1,101,006	**0.28%**
TH01	31/327,172 (0.009)	48/452,382 (0.009)	28	100/779,554	**0.01%**
TPOX	10/400,661 (0.004)	54/457,420 (0.012)	28	100/857,481	**0.01%**
VWA	184/564,398 (0.03)	1,482/873,547 (0.17)	814	2,480/1,437,945	**0.17%**
D3S1358	60/405,452 (0.015)	713/558,836 (0.13)	379	1,152/964,288	**0.12%**
D5S818	111/431,736 (0.025)	763/655,693 (0.12)	385	1,259/1,107,339	**0.11%**
D7S820	59/440,562 (0.013)	743/644,743 (0.12)	283	1,089/1,085,305	**0.10%**
D8S1179	96/409,869 (0.02)	779/489,968 (0.16)	364	1,239/899,837	**0.14%**
D13S317	192/482,136 (0.04)	881/621,146 (0.14)	485	1,558/1,103,282	**0.14%**
D16S539	120/467,774 (0.03)	540/484,465 (0.11)	372	1,041/952,239	**0.11%**
D18S51	186/296,244 (0.06)	1,094/494,096 (0.22)	466	1,746/790,342	**0.22%**
D21S11	464/435,388 (0.11)	772/526,708 (0.15)	580	1,816/962,096	**0.19%**
Penta D	12/18,701 (0.06)	21/22,301 (0.09)	24	57/41,202	**0.14%**
Penta E	20/44,311 (0.065)	75/55,719 (0.135)	59	163/100,030	**0.16%**
D2S1338	15/72,830 (0.021)	137/152,310 (0.10)	90	262/225,140	**0.12%**
D19S433	38/70,001 (0.03)	78/103,489 (0.075)	71	187/173,490	**0.11%**
SE33 (ACTBP2)	0/330 (<0.30)	330/51,610 (0.64)	None reported	330/51,940	**0.64%**

*Data used with permission from American Association of Blood Banks (AABB) 2003 Annual Report

Paternal = 0.32%

Maternal = 0.05%

6.4-fold increase in males

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

Value of STR Kits

Advantages

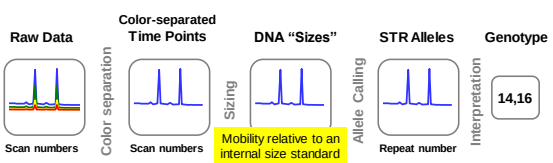
- Quality control of materials is in the hands of the manufacturer (saves time for the end-user)
- Improves consistency in results across laboratories – same allelic ladders used
- Common loci and PCR conditions used – aids DNA databasing efforts
- Simpler for the user to obtain results

Disadvantages

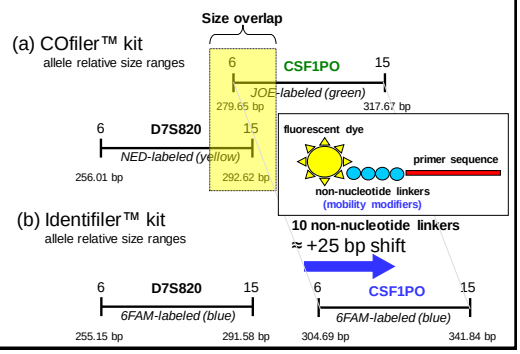
- Contents may not be completely known to the user (e.g., primer sequences)
- Higher cost to obtain results

What is Being Measured with STR Alleles during CE Separation

- Mobility of a PCR product with a fluorescent tag is being measured
- Mobility is the time it takes for the DNA molecule to move from the injection point to the detection point
- Mobility modifiers are used in some ABI STR kits
  - Identifiler has five loci with mobility modifiers



Mobility Modifiers Permit Shifting Allele Sizes without Changing Primer Binding Sites



To Avoid Overlapping PCR Product Size Ranges with STR Loci in the Same Dye Channel

- Applied Biosystems Strategy
  - **Maintains primer sequences** (except MiniFiler & NGM kits)
  - Utilizes mobility modifiers or additional dyes, no primer redesign is necessary
  - Enables comparison to legacy data with earlier kits but null alleles may go undetected with the potential for incorrect genotypes within data sets
- Promega Corporation Strategy
  - Moves primer sequences to change PCR product size ranges
  - Primer redesign can be difficult, but can be moved from primer-binding-site mutations
  - **Requires concordance studies to check for potential allele dropout**

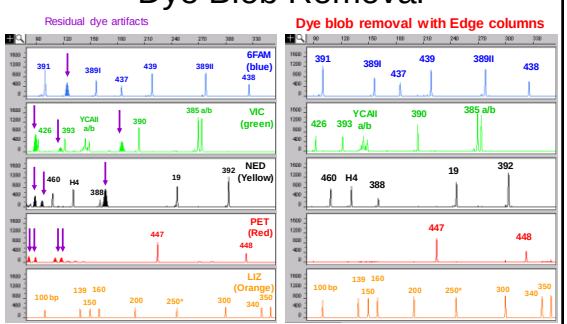
Forensic DNA Issues

- Dye Blobs
- Pull up
- Degraded DNA
- PCR Inhibition
- Contamination
- Mixed samples

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)

Dye Blob Removal

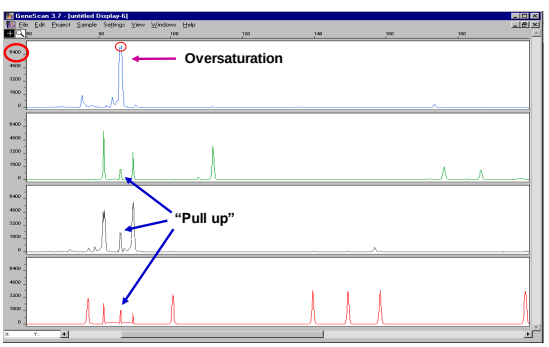


NIST Y-STR 20plex assay

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

PCR Artifacts – Pull Up

With degraded DNA two injections may be necessary to keep data on-scale



(Review)

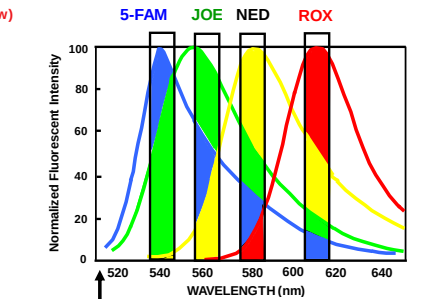
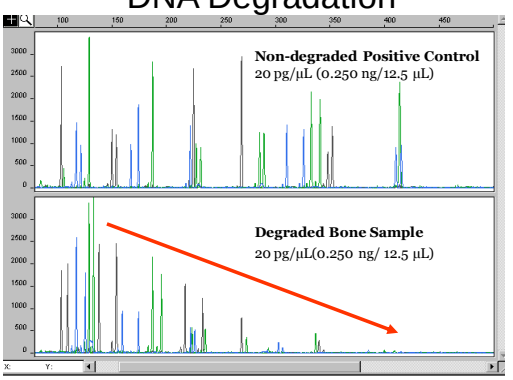


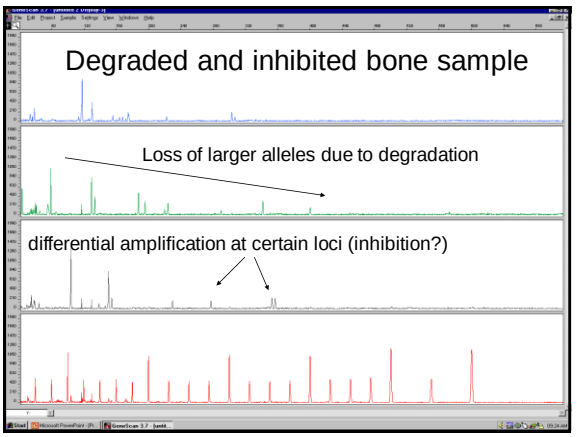
Figure 13.4, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Degradation and PCR Inhibition

- Degradation affects larger alleles more, however there is no published study on the "threshold at which degradation is apparent"
  - The amplification efficiency of each set of alleles varies independently and differential amplification across loci can occur – Moretti, JFS 2001
  - Low quality formamide can mimic the degradation effect
  - Inhibition generally affects certain loci more than others and may or may not produce a slope effect- McCord, unpublished
  - There are several likely mechanisms for inhibition including DNA aggregation, Protein-DNA binding, chelation of Mg, interference with primer binding, etc.

DNA Degradation





Non-DNA Contamination/Inhibition

- Anything that is water soluble may co-extract with DNA unless a capture technique is used.
- For capture techniques anything with a similar chemical property to DNA may co-extract
- Detergents, metal ions, humic substances are all potent contaminant/inhibitors
- Can cause all sorts of strange effects including
  - Spikes, dye blobs, elevated baselines, loss of signal, odd current effects

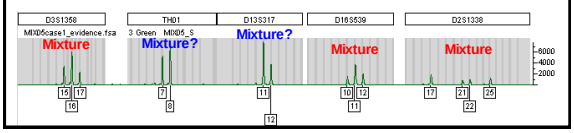
Inhibition vs. Degradation

- Will often give a similar profile.
- If a sample is inhibited – diluting the sample can often increase PCR amplification success.
- For a degraded sample, potentially use miniSTRs or LCN.

Contamination/Mixtures: Issues and Challenges

From J.M. Butler (2005) Forensic DNA Typing, 2nd Edition, p. 155

- The probability that a mixture will be detected improves with the use of more loci and genetic markers that have a high incidence of heterozygotes.
- The detectability of multiple DNA sources in a single sample relates to the ratio of DNA present from each source, the specific combinations of genotypes, and the total amount of DNA amplified.
- Some mixtures will not be as easily detectable as other mixtures.



Thank you for your attention

Acknowledgments: NIJ & FBI Funding

Contact Information

Becky Hill
Research Biologist
becky.hill@nist.gov
301-975-4275

Mike Coble
Research Biologist
michael.coble@nist.gov
301-975-4330

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