STRs, CE, and Mixtures

Florida Statewide DNA Training

Indian Rocks Beach (Largo), FL

May 12-13, 2008

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

john.butler@nist.gov

Topics and Techniques for Forensic DNA Analysis

Introductions

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

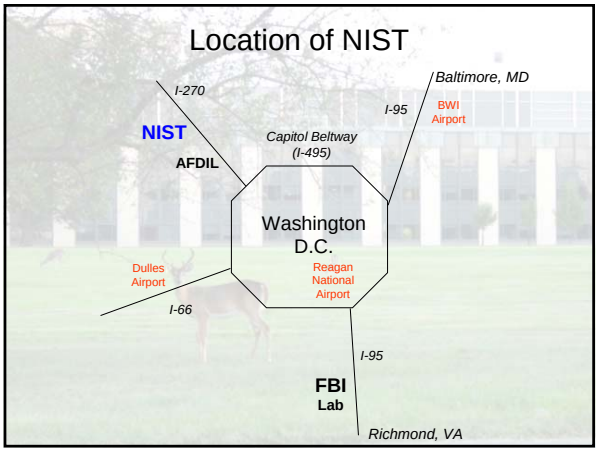
NIST History and Mission

- National Institute of Standards and Technology (NIST) was created in 1901 as the National Bureau of Standards (NBS). The name was changed to NIST in 1988.
- NIST is **part of the U.S. Department of Commerce** with a mission to develop and promote measurement, standards, and technology to enhance productivity, facilitate trade, and improve the quality of life.
- NIST supplies over 1,300 Standard Reference Materials (SRMs) for industry, academia, and government **use in calibration of measurements**.
- **NIST defines time for the U.S.**

$603 for 3 jars


DNA typing standard

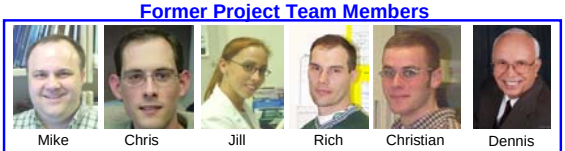
Location of NIST



NIST Human Identity Project Team

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

Former Project Team Members



Mike Coble — AFDIL
Chris DeAngelis — Medical School
Jill Appleby — NC SBI
Rich Schoske — Air Force
Christian Ruitberg — Pharma
Dennis Reeder — Retired/ABI

National Institute of Justice

The Research, Development, and Evaluation Agency of the U.S. Department of Justice

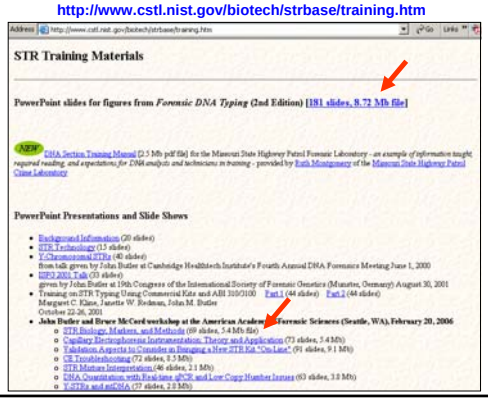
Current Areas of NIST Effort with Forensic DNA

- **Standards**
  - Standard Reference Materials
  - Standard Information Resources (STRBase website)
  - Interlaboratory Studies
- **Technology**
  - Research programs in SNPs, miniSTRs, Y-STRs, mtDNA, qPCR
  - Assay and software development, expert system review
- **Training Materials**
  - Review articles and workshops on STRs, CE, validation
  - PowerPoint and pdf files available for download

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

Training Materials Available on STRBase

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



Contributors to These Workshop Slides

Bruce McCord

Florida International University

CE

Mike Coble

AFDIL

miniSTRs

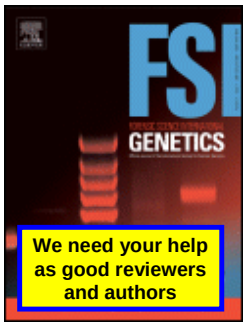
Angie Dolph

Marshall U./ NIST

mixtures

Forensic Science International: Genetics

http://www.fsigenetics.com/



Editor-in-Chief:
Angel Carracedo (Spain)
Associate Editors:
Peter M. Schneider (Germany)
John M. Butler (USA)

FSI: Genetics is a new journal dedicated exclusively to the field of forensic genetics. It has been launched in 2007 by Elsevier Publishers in affiliation with the International Society of Forensic Genetics. **All members of the ISFG receive a free subscription of this journal** (print and online version) as part of their membership benefits.

Primary Sources for Material Covered in this Workshop

- Butler, J.M., Buel, E., Crivellente, F., and McCord, B.R. (2004) Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25: 1397-1412
- Butler, J.M. (2006) Genetics and genomics of core STR loci used in human identity testing. *J. Forensic Sci.* 51(2): 253-265
- McCord, B. (2003) Troubleshooting capillary electrophoresis systems. *Profiles in DNA* 6(2): 10-12 (Promega Corporation); available at http://www.promega.com/profiles/602/ProfilesInDNA_602_10.pdf
- Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition: Biology, Technology, and Genetics of STR Markers*. Elsevier Science/Academic Press
- NIST STRBase website: http://www.cstl.nist.gov/biotech/strbase/

These workshop materials will be made available at http://www.cstl.nist.gov/biotech/strbase/training.htm

Outline for Workshop

Day 1

- STRs and Artifacts
- miniSTRs
- CE Troubleshooting

LUNCH

- Dawn Herkenham (Legal Issues)

Day 2

- Mixture Interpretation
- Mixture Examples

LUNCH

- Mixture Stats
- Interlab Studies
- Company presentations

My Goal is to Answer YOUR Questions – So Please Ask Them…

Understanding the Audience Here

- Where is everyone from?
  - State lab?
  - Local lab?
  - Private lab?
- Experience level?
  - Less than 1 year?
  - 1-3 years?
  - >3 years?
- STR kits in use?
  - Profiler Plus/COfiler
  - Identifiler
  - PowerPlex 16
  - Y-STRs?
- Instrumentation is use?
  - ABI 310
  - ABI 3100/3130xl
  - Other?
- Software in use?
  - GeneScan/Genotyper
  - GeneMapperID
  - Other?

NIST and NIJ Disclaimer

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

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

Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

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

Topics and Techniques for Forensic DNA Analysis

STRs and Molecular Biology Artifacts

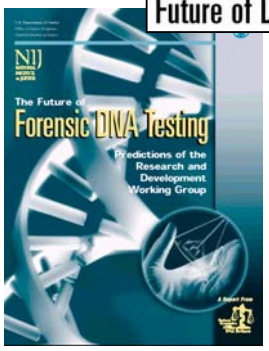
Florida Statewide Training Meeting

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

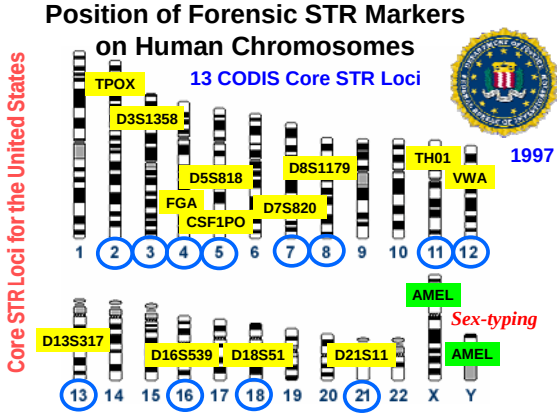
Conclusions

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

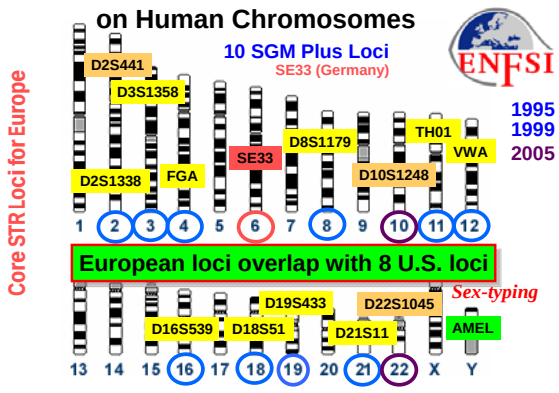
Advantages for STR Markers

- 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



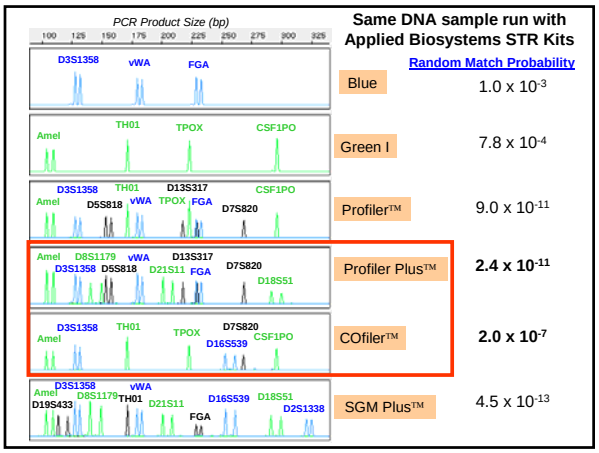
Value of STR Kits

Advantages

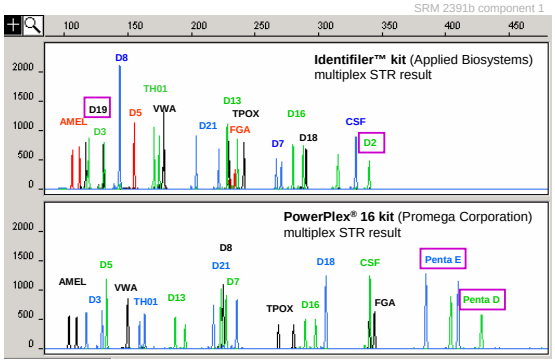
- Quality control of materials is in the hands of the manufacturer (saves time for the end-user)
- Improves consistency in results across laboratories – same allelic ladders used
- Common loci and PCR conditions used – aids DNA databasing efforts
- Simpler for the user to obtain results

Disadvantages

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



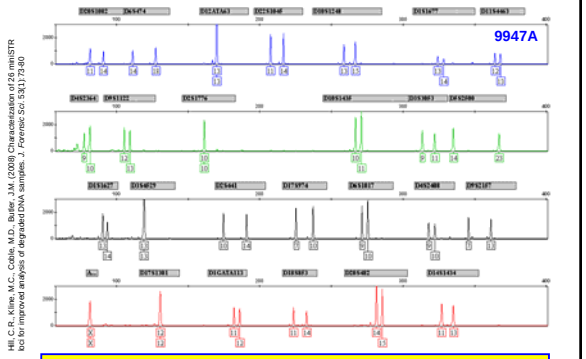
Commercial STR 16plex Kits



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

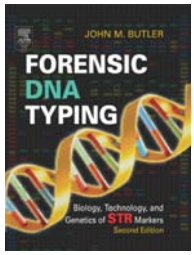
Category	Example Repeat Structure	13 CODIS Loci
Simple repeats – contain units of identical length and sequence	(GATA)(GATA)(GATA)	TPOX, CSF1PO, D5S818, D13S317, D16S539
Simple repeats with non-consensus alleles (e.g., TH01 9.3)	(GATA)(GAT-)(GATA)	TH01, D18S51, D7S820
Compound repeats – comprise two 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

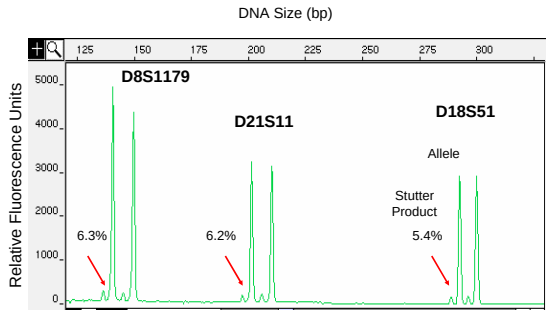
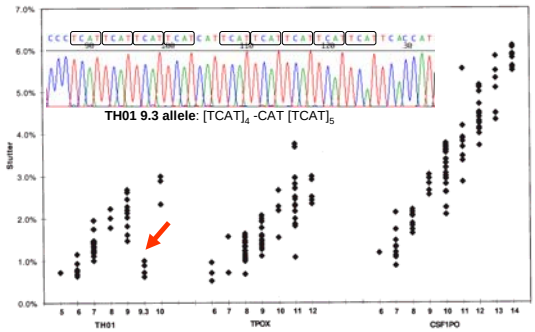


Figure 6.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Measured Stutter Percentages

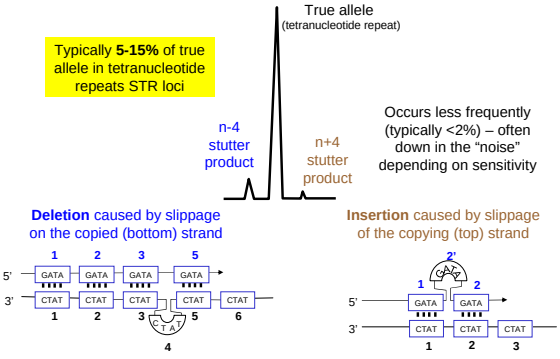
Variable by Allele Length and Composition



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.

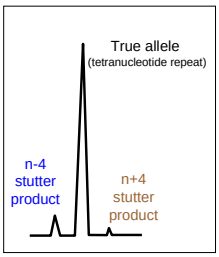
Stutter Product Formation

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**



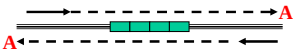
N-4 Stutter % of	main allele: allele	main allele: rfu	N+4: 'allele'	N+4: rfu	N+4 Stutter % of
6.43%	19	4664	20	57	1.22%

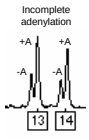
http://www.cstl.nist.gov/biotech/strbase/validation/N+4_stutter_spreadsheet.xls

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)

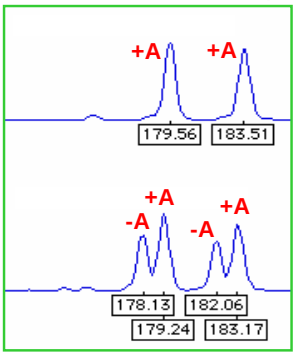
Best if there is NOT a mixture of "+/- A" peaks
(desirable to have full adenylation to avoid split peaks)





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

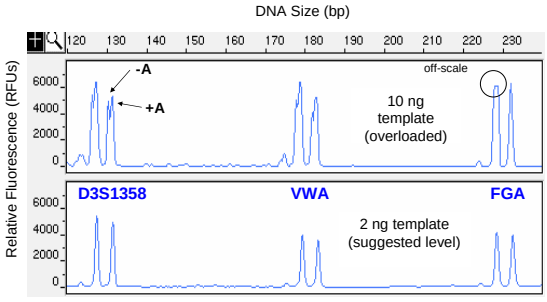
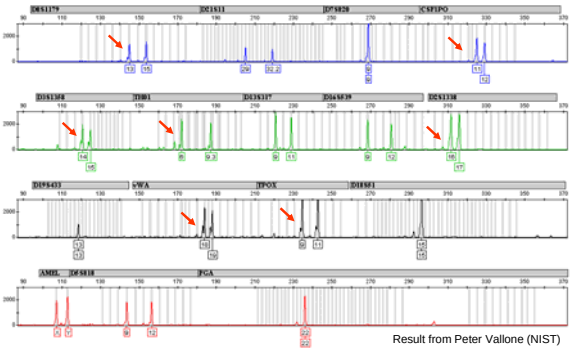


Figure 6.5, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Identifiler – Rapid PCR (36 min total time)

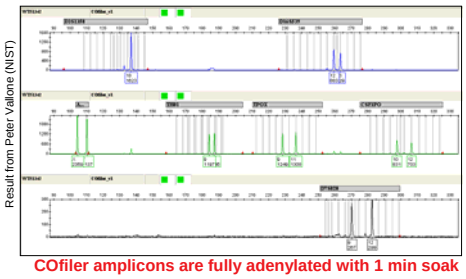
with 1 min 60 ºC adenylation soak (using different polymerases)



Result from Peter Vallone (NIST)

Rapid PCR Work and Adenylation

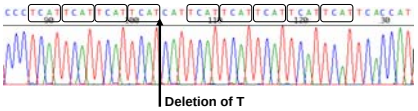
- **Poor adenylation (presence of –A peaks) is locus-specific and impacted by number of loci amplified**



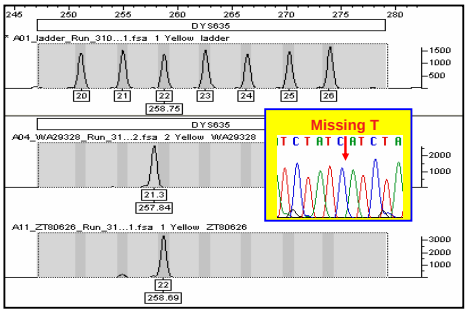
COfiler amplicons are fully adenylated with 1 min soak

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

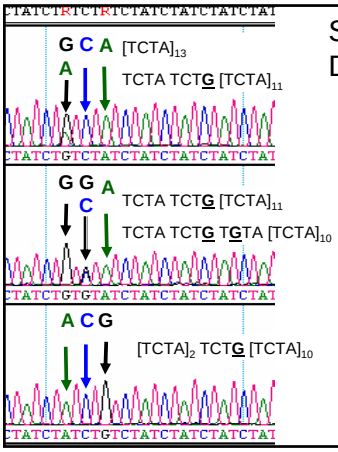


Allele 22 bin
258.75 +/- 0.5 = 258.25 to 259.25

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.

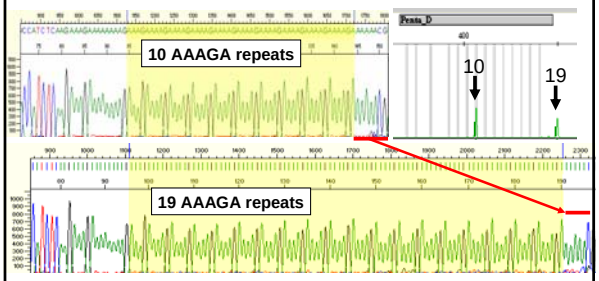
We require a minimum of 10 ng for the sequencing.
We request copies of the electropherograms demonstrating the variant allele.
The more information we have up front the better.
Please have patience we will get to your samples!

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)!

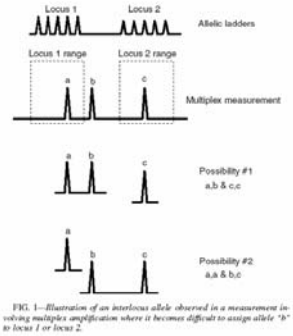
- Contact Margaret Kline: **margaret.kline@nist.gov**

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

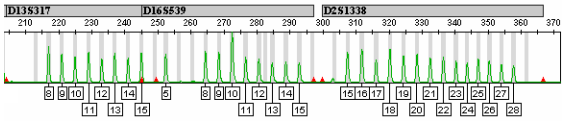


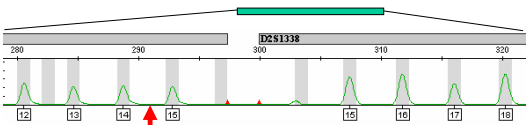
- Use a different multiplex STR kit with different locus combinations
- Test singleplex for each putative locus
- Example: Identifiler D16S539 and D2S1338

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)

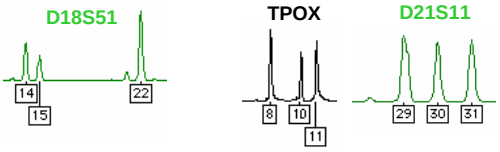
Likely a D2S1338 allele 11

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

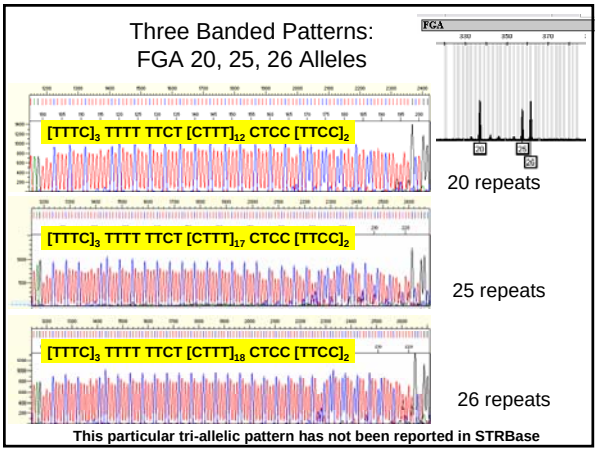
Most common in D18S51 and …..

"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

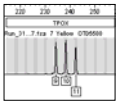
A.B. Lane

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%

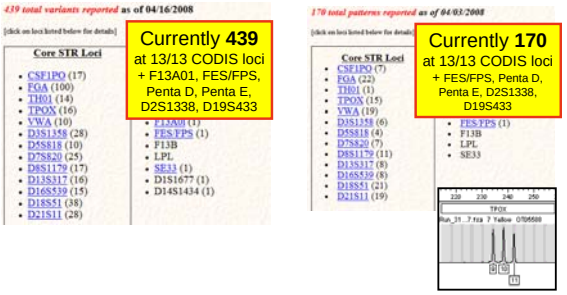
In 78 observations of 16 different TPOX tri-allelic patterns, only 4 times (5%) is allele "10" not present

Variant Alleles Cataloged in STRBase

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

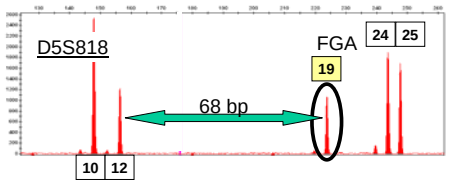
Off-Ladder Alleles

Tri-Allelic Patterns



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

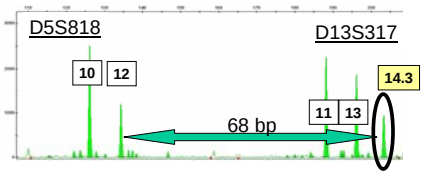
PK HT Ratio
12/10 - 0.48



PK HT Ratio
19/24 - 0.55
25/24 – 0.89

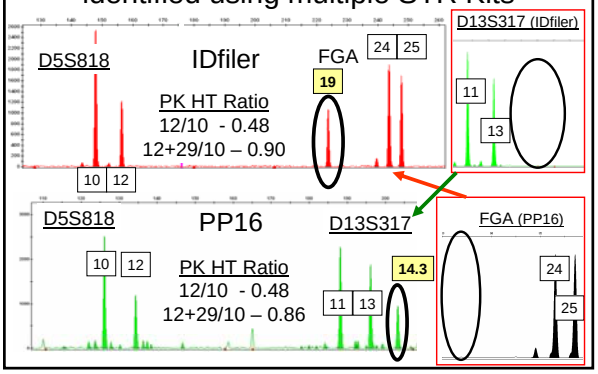
Or is this a D13S317 - Tri-allelic pattern identified using Powerplex 16?

PK HT Ratio
12/10 - 0.48

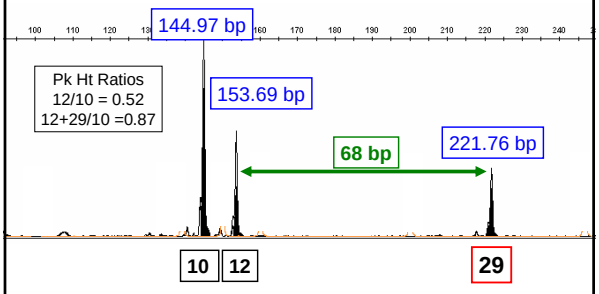


PK HT Ratio
13/11 - 0.83
14.3/11 – 0.42

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

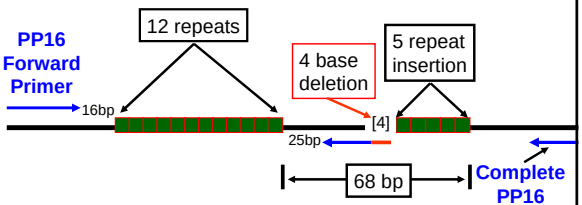


D5S818 monoplex results



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?

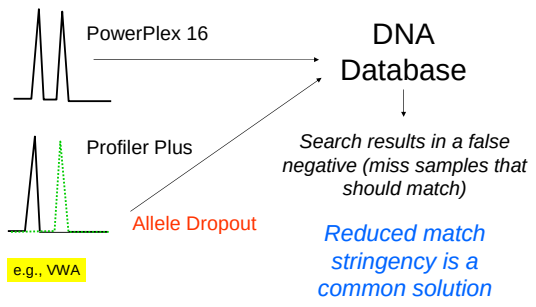
- STRBase Tri-allelic reports for FGA for 19,*,* patterns with AB amplification kits.
  - 5 reports :
  - 19,20,21; 19,20,23; 19,20,24; 19,22,23; 19,24,25
  - But there we have sequenced true tri-allelic FGA samples
- STRBase Tri-allelic reports for D13S317 for *,*, OL patterns with PP16 amplification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

Null Alleles

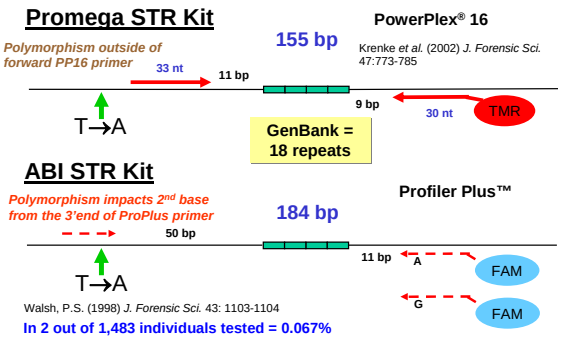
- Allele is present in the DNA sample but fails to be amplified due to a **nucleotide change in a primer binding site**
- Allele dropout is a problem because a heterozygous sample appears falsely as a homozygote
- Two PCR primer sets can yield different results on samples originating from the same source
- 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



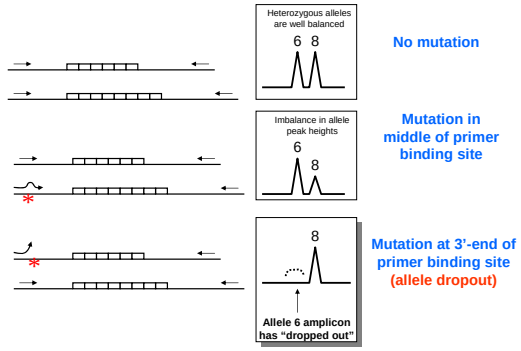
Krenke *et al.* (2002) *J. Forensic Sci.* 47:773-785

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

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

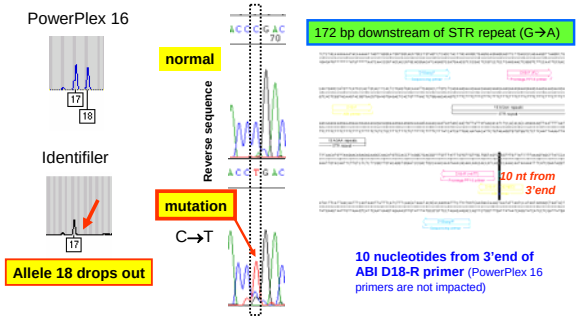
Lazaruk *et al.* (2001) *Forensic Sci Int.* 119:1-10

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

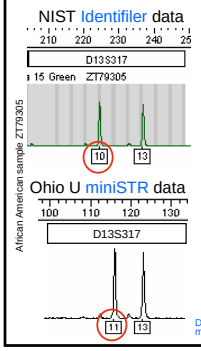
D18S51 Null Allele from Kuwait Samples with ABI Primers



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

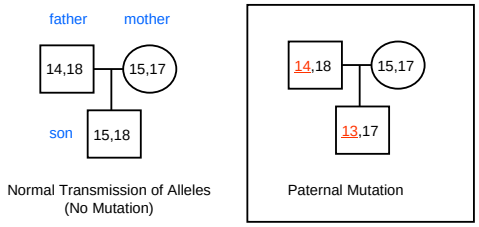
10/13 CODIS loci affected so far

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

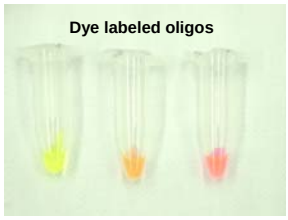
Mutations impact paternity testing and missing persons investigations but not forensic direct evidence-suspect matches…

- 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

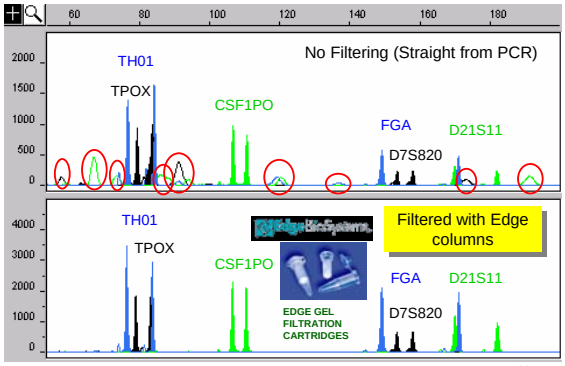


6FAM (yellow), **VIC** (orange), **NED** (red)

- 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)

Topics and Techniques for Forensic DNA Analysis

miniSTRs

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 Institute of Justice
The Research, Development, and Evaluation Agency of the U.S. Department of Justice

Current Areas of NIST Effort with Forensic DNA

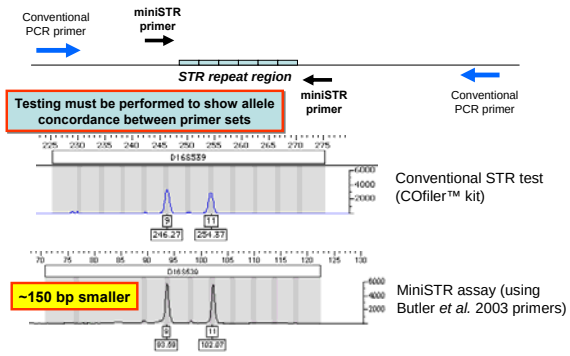
- **Standards**
  - Standard Reference Materials
  - Standard Information Resources (STRBase website)
  - Interlaboratory Studies
- **Technology**
  - Research programs in SNPs, miniSTRs, Y-STRs, mtDNA, qPCR
  - Assay and software development, expert system review
- **Training Materials**
  - Review articles and workshops on STRs, CE, validation
  - PowerPoint and pdf files available for download

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

Technology: Research Programs

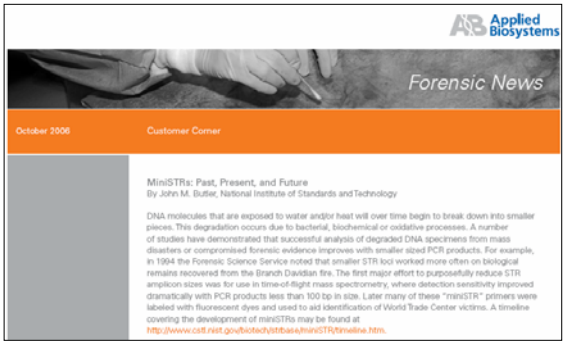
- miniSTRs
- Y-chromosome STRs
- mtDNA
- SNPs
- qPCR for DNA quantitation
- DNA stability studies
- Variant allele characterization and sequencing
- Software tools
- Expert System review
- Assay development with collaborators

A miniSTR is a reduced size STR amplicon that enables higher recovery of information from degraded DNA samples



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 7.2, ©Elsevier Science/Academic Press

miniSTR Overview Article



http://marketing.appliedbiosystems.com/images/enews/ForensicNews_Vol7/PDF/02A_CustomerCorner_Butler.pdf

Timeline for miniSTRs

and Demonstrating the Value of Using Reduced Size Amplicons for Degraded DNA

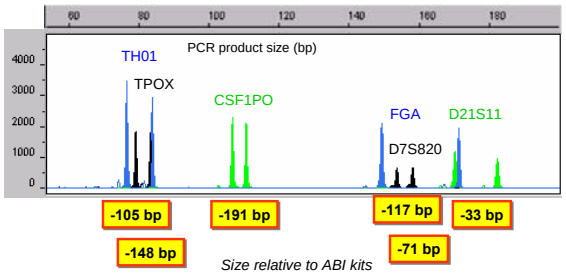
- 1994 – FSS finds that smaller STR loci work best with burned bone and tissue from Branch Davidian fire
- 1997 – New primers developed for time-of-flight mass spectrometry to make small STR amplicons
- 2001 – Work at NIST and OhioU with CODIS STRs; BodePlexes used in WTC investigation starting 2002
- 2004 – Work at NIST with **non-CODIS (NC) miniSTRs**
- 2007 – Applied Biosystems releases 9plex MiniFiler

http://www.cstl.nist.gov/biotech/strbase/miniSTR/timeline.htm

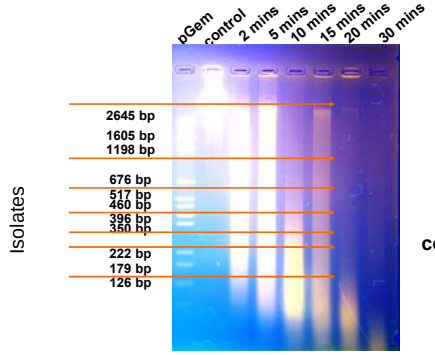
J. Forensic Sci. Sept 2003 issue

John M. Butler,[1] Ph.D.; Yin Shen,[2,3] Ph.D.; and Bruce R. McCord Ph.D.[2]

The Development of Reduced Size STR Amplicons as Tools for Analysis of Degraded DNA*

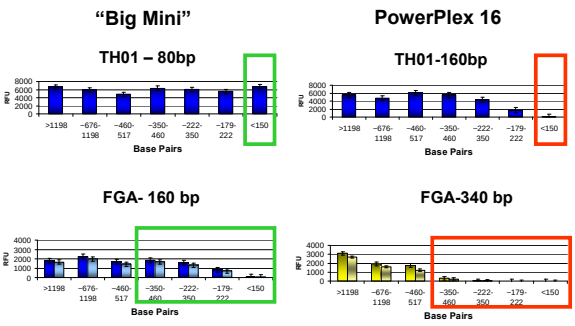


DNA Degraded With DNase I



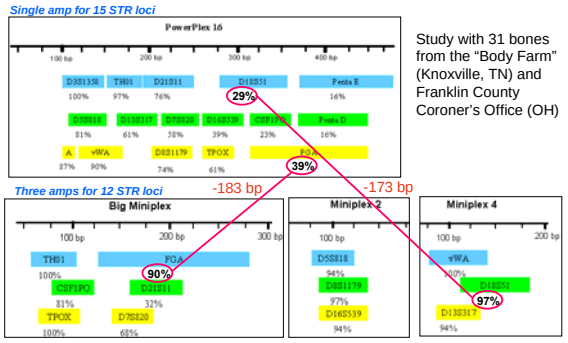
Chung, D.T., Drabek, J., Opel, K.L., Butler, J.M., McCord, B.R. (2004) A study on the effects of degradation and template concentration on the efficiency of the STR miniplex primer sets. *J. Forensic Sci.* 49(4): 733-740.

Miniplexes improve detection of degraded DNA



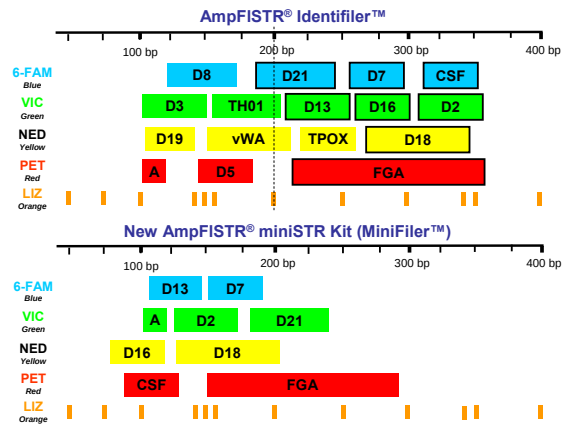
Chung, D.T., Drabek, J., Opel, K.L., Butler, J.M., McCord, B.R. (2004) A study on the effects of degradation and template concentration on the efficiency of the STR miniplex primer sets. *J. Forensic Sci.* 49(4): 733-740.

Comparison of PCR Amplification Success Rates with Commercial Kit vs. miniSTR Assays



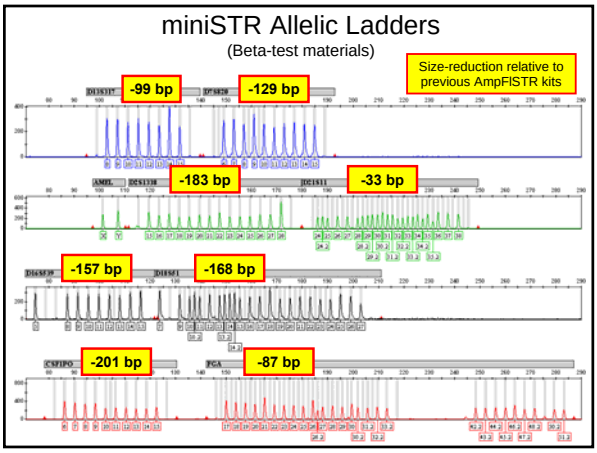
Study with 31 bones from the "Body Farm" (Knoxville, TN) and Franklin County Coroner's Office (OH)

Opel K. L.; Chung, D. T.; Drábek, J.; Tatarek, N. E.; Jantz, L. M.;. McCord, B.R. (2006) The Application of Miniplex Primer Sets in the Analysis of Degraded DNA from Human Skeletal Remains. *J. Forensic Sci.* 51(2): 351-356.



miniSTR Allelic Ladders

(Beta-test materials)



Summary of Samples Typed with ABI MiniFiler kit at NIST and ABI

- Primarily only population samples examined – no extensive sensitivity or degraded DNA tests were performed

1,308 samples Allele concordance = 10,437/10,464 = 99.7%

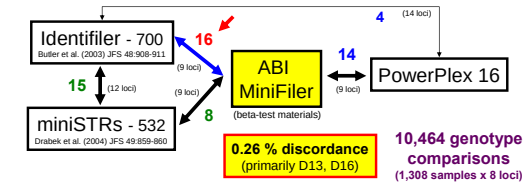
- **656 NIST U.S. population samples**
  - 260 Caucasian, 253 African American, 140 Hispanic, 3 Asian
  - Previously examined with **Identifiler; also with PowerPlex 16**
  - Also tested with Butler *et al.* (2003) **published miniSTR primers**
  - http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm
- **481 father-son pairs**
  - 184 Caucasian, 196 African American, 101 Asian samples (provided by paternity testing company DDC)
  - Previously examined with **Identifiler**
- **171 samples from Applied Biosystems**

Hill *et al.* (2007) Concordance study between the AmpFlSTR MiniFiler PCR Amplification Kit and conventional STR typing kits. *J. Forensic Sci.* 52(4): 870-873.

Concordance Conducted at NIST

27 Discordant Calls

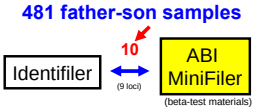
656 NIST U.S. population samples



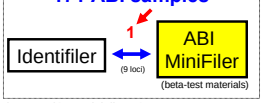
0.26 % discordance (primarily D13, D16)

10,464 genotype comparisons (1,308 samples x 8 loci)

481 father-son samples



171 ABI samples

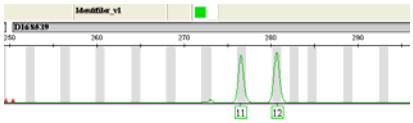


Hill *et al.* (2007) Concordance study between the AmpFlSTR MiniFiler PCR Amplification Kit and conventional STR typing kits. *J. Forensic Sci.* 52(4): 870-873.

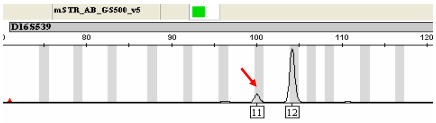
Concordance Studies Reveal Potential Primer Binding Site Mutations with Different Primer Sets

D16S539

Identifiler



miniSTR Kit (beta-test)



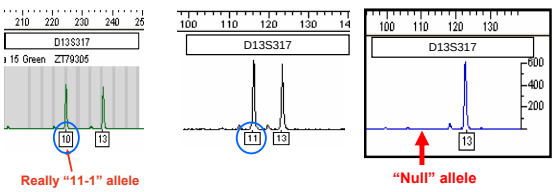
Appears to be an allele 11 dropout/reduction due to primer binding site mutation

Examination of **D13S317** Concordance:

African American sample ZT79305

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.

NIST Identifiler data | Ohio U miniSTR data | **AB miniSTR beta-test**

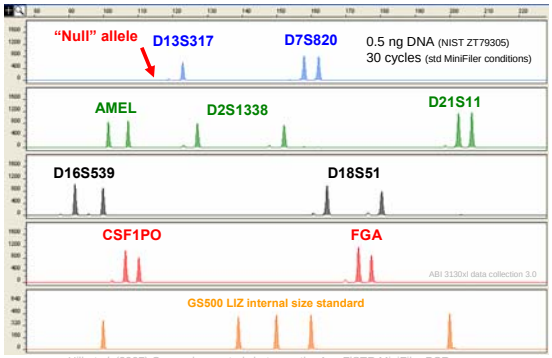


10,13 **11,13** **13,13**

Reverse primer is outside deletion | **Reverse primer is inside deletion** | **Reverse primer is on top of deletion**

Hill *et al.* (2007) Concordance study between the AmpFlSTR MiniFiler PCR Amplification Kit and conventional STR typing kits. *J. Forensic Sci.* 52(4): 870-873.

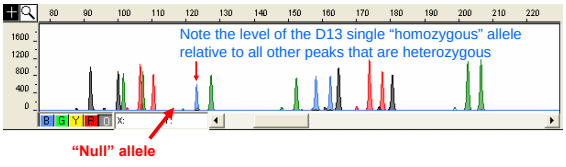
Full MiniFiler Profile for NIST Sample with D13S317 Allele Dropout

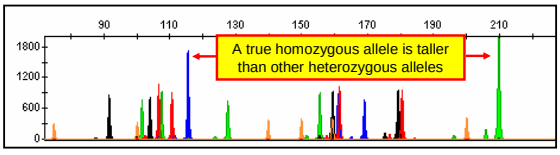


Hill *et al.* (2007) Concordance study between the AmpFlSTR MiniFiler PCR Amplification Kit and conventional STR typing kits. *J. Forensic Sci.* 52(4): 870-873.

Note the Relative D13 Peak Heights

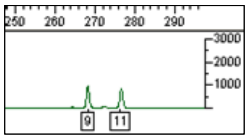
(Suggests Allele Dropout)



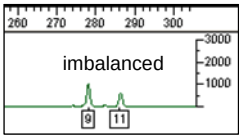


D16S539 SRM 2391b Genomic 8

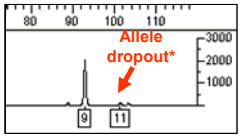
Identifiler



PowerPlex 16



MiniFiler



Due to primer binding site mutation

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

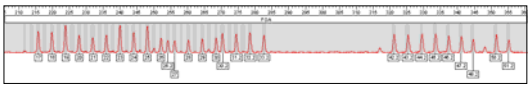
More Loci are Useful in Situations Involving Relatives

- **Missing Persons** and Disaster Victim Identification (kinship analysis)
- Immigration Testing (often limited references)
  - Recommendations for 25 STR loci
- Deficient Parentage Testing
  - often needed if only one parent and child are tested

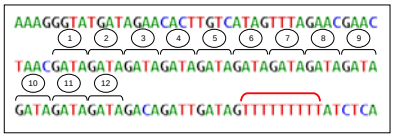
Relationship testing labs are being pushed to answer more difficult genetic questions…and **we want to make sure the right tools are in place**

Why Go Beyond the CODIS Loci?

(1) Large Allele Ranges (e.g. FGA)



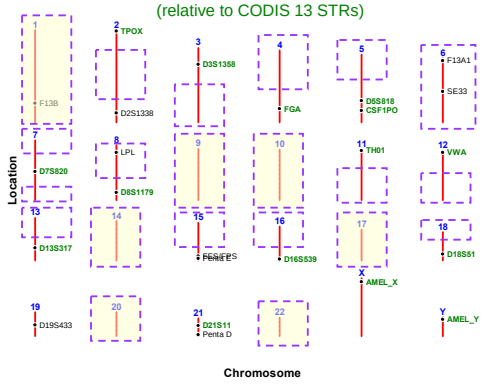
(2) "Unclean" Flanking Sequences (e.g. D7S820)

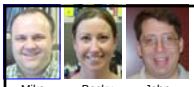


Butler, JM, Shen, Y., McCord, BR (2003) *JFS* 48(5): 1054-1064

Locations of Focus for New miniSTR Loci

(relative to CODIS 13 STRs)





New miniSTR Non-CODIS (NC) Loci

No longer at NIST (AFDIL Research Section Chief since April 2006)

- 32 STR loci tested on NIST **665 U.S. population samples**
- **26 STR loci with allele sizes below 140 bp and good heterozygosities (above TPOX level)**
- All new STR loci are **physically unlinked** to the 13 CODIS core loci
- **Submitted articles** regarding primer sequences and locus characterization including population statistics
- **SRM 2391b components are being certified** through sequencing for D10S1248, D2S441, D22S1045; for reference purposes, genotypes for standard samples (9947A, 9948, 007, K562) will be made available on STRBase

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

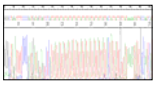
Characterization of New miniSTR Loci

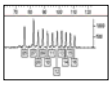
"Computer Work"

Candidate STR marker selection (e.g. Marshfield Clinic Center of Medical Genetics) → Pull down sequence data from the web (e.g. NCBI) → Identify Chromosome Location (e.g. Human BLAT Search) → Screen for PCR Primers (e.g. Primer3) → Test primers for Multiplex-ability (e.g. AutoDimer - NIST)

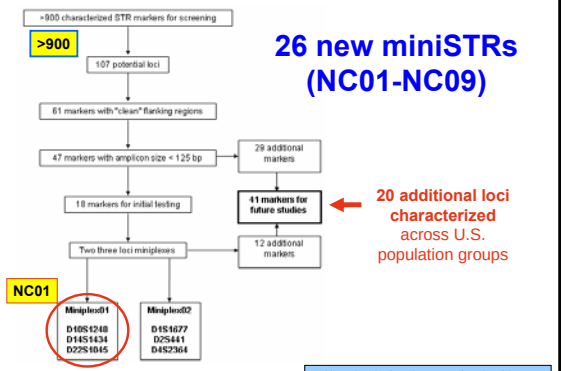
"Laboratory Work"

Test Markers on Population samples → Sequence homozygotes to determine allele sizes → Build Macros for Genotyping → Construct Allelic Ladders



Initial Testing Results with Potential miniSTR Loci



26 new miniSTRs (NC01-NC09)

20 additional loci characterized across U.S. population groups

Coble and Butler (2005) *J. Forensic Sci.* 50(1): 43-53

New STR Loci Characterized

Hill *et al.* (2008) *J. Forensic Sci.* 53(1):73-80

J Forensic Sci, January 2008, Vol. 53, No. 1
doi: 10.1111/j.1556-4029.2008.00595.x
Available online at: www.blackwell-synergy.com

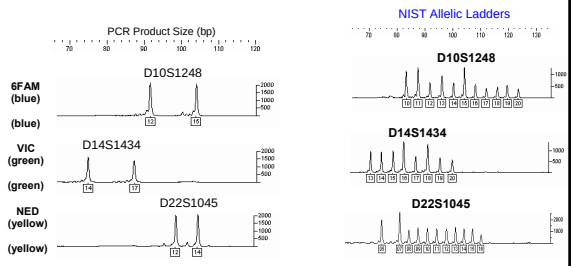
Carolyn R. Hill, M.S.; Margaret C. Kline, M.S.; Michael D. Coble,† Ph.D.; and John M. Butler, Ph.D.

Characterization of 26 MiniSTR Loci for Improved Analysis of Degraded DNA Samples

- Primer sequences (for miniplexes), GeneMapper bins and panels, genotypes on common samples, and allele frequency information **available on STRBase**

http://www.cstl.nist.gov/biotech/strbase/miniSTR.htm
http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_NC_loci_types.htm
http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_Panels_Panels.txt
http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_Panels_NC_bins_bins.txt

Miniplex "NC01"



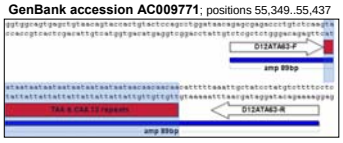
http://www.cstl.nist.gov/biotech/strbase/miniSTR.htm

Coble and Butler (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA *J. Forensic Sci.* 50(1): 43-53

Characterization of miniSTR **D12ATA63**

GenBank accession AC009771; positions 55,349..55,437

[FAM] – GAGCGAGACCCTGTCTCAAG
GAAAAGACATAGGATAGCAATTT



Chr 12 106.825 Mb (12q23.3)

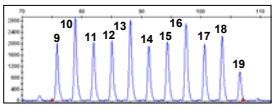
Trinucleotide [TAA][CAA] repeat
76 -106 bp
Alleles 9 -19

Allele	Caucasian (N = 260)	African Am (N = 259)	Hispanic (N = 140)
9	--	--	0.0036
10	0.0019	0.0154	0.0036
11	0.1385	0.1525	0.1500
12	0.2154	0.1004	0.1786
13	0.0173	0.1564	0.0286
14	0.1615	**0.3340**	0.2214
15	0.0577	0.0772	0.0714
16	**0.2981**	0.1004	**0.2643**
17	0.0981	0.0521	0.0679
18	0.0096	0.0058	0.0071
19	0.0019	0.0058	0.0036

Heterozygosity Values

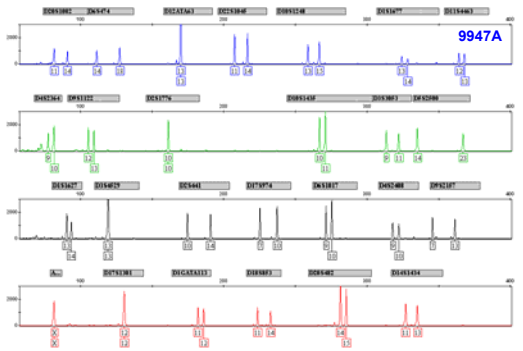
U.S. Caucasian	**0.842**
African American	**0.788**
U.S. Hispanic	**0.879**

D12ATA63 Allelic Ladder



"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

European Labs Have Adopted the NIST-Developed NC miniSTRs

FSI (2006) **156(2):** 242-244

Short communication

The evolution of DNA databases—Recommendations for new European STR loci

Peter Gill [a,*], Lyn Fereday [b], Niels Morling [c], Peter M. Schneider [d]

[a] Forensic Science Service, Birmingham, UK
[b] Forensic Science Service, London, UK
[c] Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Denmark
[d] Institute of Legal Medicine, University of Cologne, Germany

Received 25 May 2005; accepted 26 May 2005

…recommended that existing multiplexes are re-engineered to enable small amplicon detection, and that three new mini-STR loci with alleles <130 bp (D10S1248, D14S1434 and D22S1045) are adopted as universal. This will increase the number of European standard Interpol loci from 7 to 10.

(D14 has been replaced with D2S441 from NC02)

Summary of miniSTRs

- **Reduced size amplicons improve success rates with degraded DNA** or samples possessing PCR-inhibitors – European leaders view **miniSTRs as "the way forward"**
- **MiniFiler concordance** testing performed
- **New miniSTR loci are being characterized** at NIST – 26 loci developed

Thank you for your attention…

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

Becky Hill

miniSTRs and 26plex work

Bruce McCord

Early miniSTR work

Mike Coble

Original NC miniSTR work

Margaret Kline

STR allele sequencing

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

Collaborators from ABI
Lori Hennessy
Julio Mulero
Rob Lagace
Chien-Wei Chang

Topics and Techniques for Forensic DNA Analysis

Capillary Electrophoresis Fundamentals and Troubleshooting

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

Questions?

- What are your biggest challenges with keeping your ABI 310/3100/3130xl running?
- What kind of signal intensity variation are you seeing between your different instruments?
- Have anyone seen uneven injection across a sample plate? (We believe this to be an autosampler calibration issue…e.g., position G10 or H12 does not inject properly)

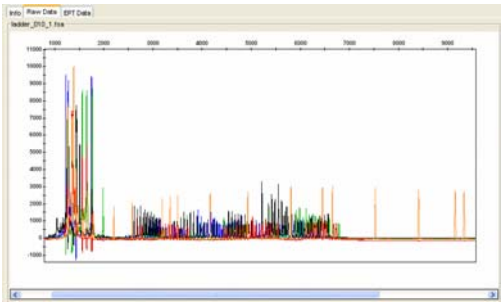
Planned Promega 2008 Meeting Troubleshooting Workshop

- Title: **"Principles of Interpretation and Troubleshooting of Forensic DNA Typing Systems"**
- Instructors: **John Butler (NIST) and Bruce McCord (FIU)**
- Date: **October 16, 2008** with Promega Int. Symp. Human ID

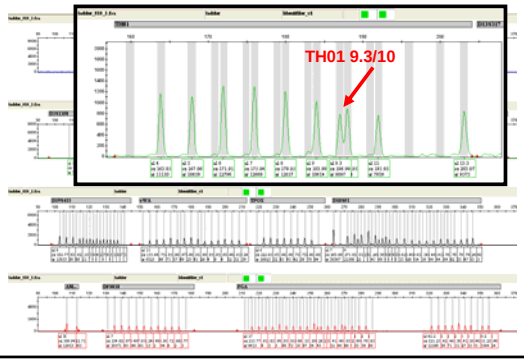
The workshop will consist of three parts:
(1) a through examination of theoretical issues with capillary electrophoresis PCR amplification of short tandem repeat markers
(2) a discussion of how to properly set instrument parameters to interpret data (including mixtures), and
(3) **a review of specific problems seen by labs** submitting problematic data and commentary on possible troubleshooting solutions.

Seeking input of problems observed with CE systems

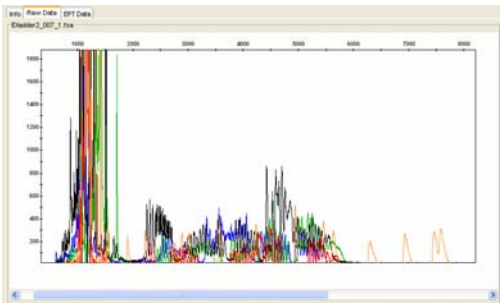
Identifiler Allelic Ladder March 14, 2007



Identifiler Allelic Ladder March 14, 2007

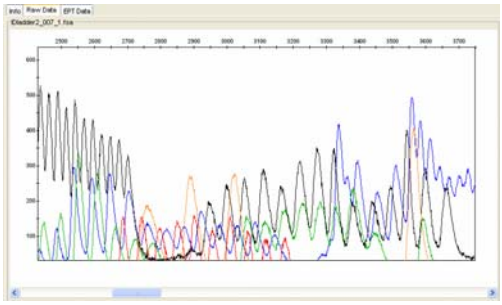


Identifiler Allelic Ladder March 23, 2007

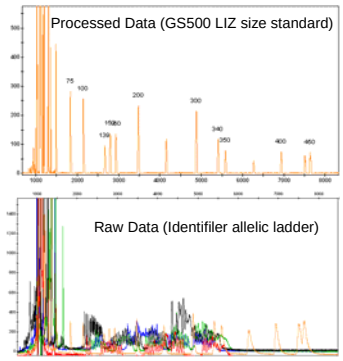


Same capillary array, same POP6 polymer, …

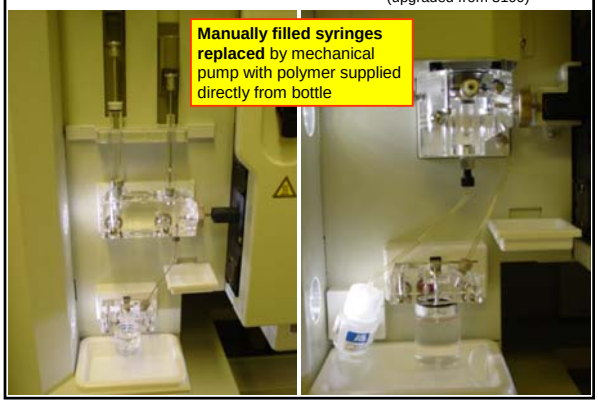
Examination of Resolution in TH01 Region



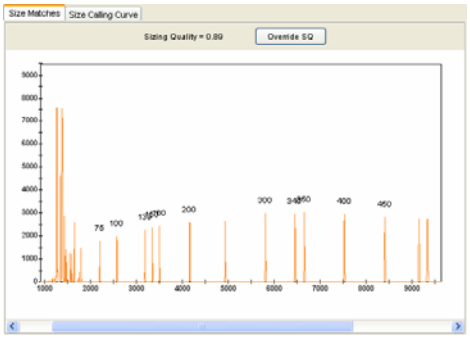
Examine the Size Standard…



ABI 3100 / ABI 3130xl (upgraded from 3100)



The Size Standard Provides an Excellent Indicator of Performance on Every Sample



Review Article on STRs and CE

pdf available from http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

Electrophoresis 2004, *25*, 1397–1412

Review

John M. Butler[1]
Eric Buel[2]
Federica Crivellente[3]*
Bruce R. McCord[3]

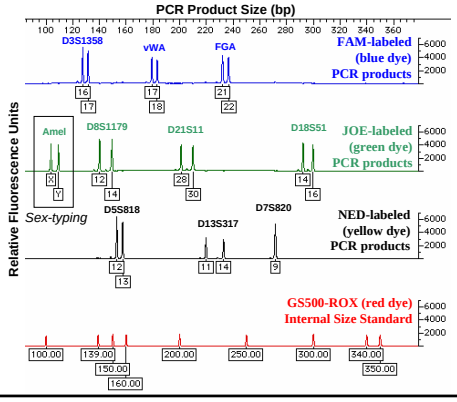
[1]National Institute of Standards and Technology, Biotechnology Division, Gaithersburg, MD, USA
[2]Vermont Forensic Laboratory, Waterbury, VT, USA
[3]Ohio University, Department of Chemistry, Athens, OH, USA

Forensic DN
using the AB
for STR anal

DNA typing with shor
applications includin
such as the ABI Prism
for many laboratories
ing sample preparat
results using CE syst
ored in the context
throughput and ease

Contents

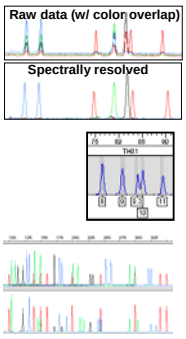
Genotype Results with Profiler Plus™ kit



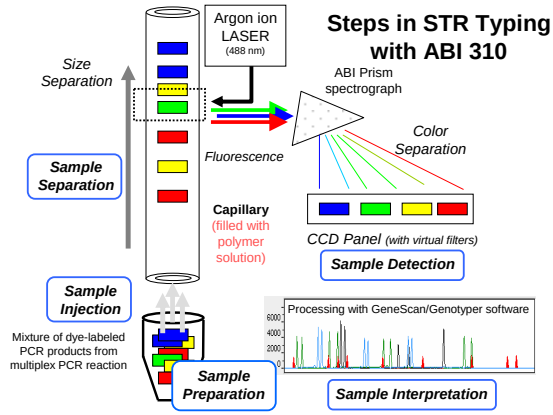
Analytical Requirements for STR Typing

Butler *et al.* (2004) Electrophoresis 25: 1397-1412

- Fluorescent dyes must be **spectrally resolved** in order to distinguish different dye labels on PCR products
- PCR products must be **spatially resolved** – desirable to have single base resolution out to >350 bp in order to distinguish variant alleles
- High **run-to-run precision** – an internal sizing standard is used to calibrate each run in order to compare data over time



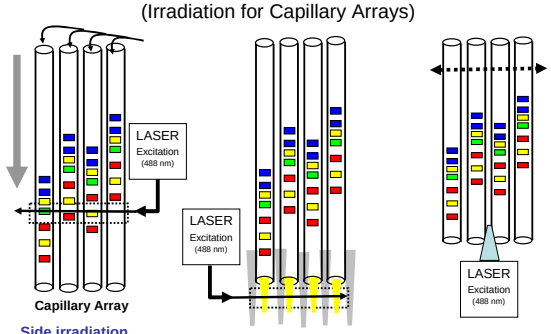
Steps in STR Typing with ABI 310



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 13.8, © Elsevier Science/Academic Press

Detection with Multiple Capillaries

(Irradiation for Capillary Arrays)



Process Involved in 310/3100 Analysis

- Separation
  - Capillary – 50um fused silica, 43 cm length (36 cm to detector)
  - POP-4 polymer – Polydimethyl acrylamide
  - Buffer - TAPS pH 8.0
  - Denaturants – urea, pyrolidinone
- Injection
  - electrokinetic injection process (formamide, water)
  - importance of sample stacking
- Detection
  - fluorescent dyes with excitation and emission traits
  - CCD with defined virtual filters produced by assigning certain pixels

Ohm's Law

- V = IR (where V is voltage, I is current, and R is resistance)
- Current, or the flow of ions, is what matters most in electrophoresis
- CE currents are much lower than gels because of a higher resistance in the narrow capillary
- CE can run a higher voltage because the capillary offers a higher surface area-to-volume ratio and can thus dissipate heat better from the ion flow (current)

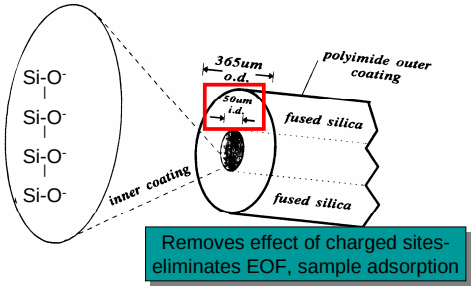
Separation Issues

- Electrophoresis buffer –
  - Urea for denaturing and viscosity
  - Buffer for consistent pH
  - Pyrolidinone for denaturing DNA
  - EDTA for stability and chelating metals
- Polymer solution -- POP-4 (but others work also)
- Capillary wall coating -- dynamic coating with polymer
  - Wall charges are masked by methyl acrylamide
- Run temperature -- 60 ºC helps reduce secondary structure on DNA and improves precision. (Temperature control affects DNA sizing)

Why TAPS instead of Tris-borate (TBE) buffer?

- TBE is temperature/pH sensitive
  - as temperature increases, the pH decreases (0.02 pH units with every 1 °C); this is the principle by which TaqGold activation works
- At lower pH, fluorescence emission of dyes decreases
  - see Singer and Johnson (1997) *Proceedings of the Eighth International Symposium on Human Identification*, pp. 70-77
- Thus when running at 60 °C on the ABI 310, if Tris-borate was used, fluorescent intensity of PCR products would be lower

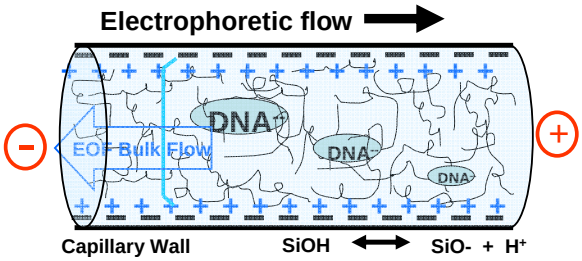
Capillary Coating



Removes effect of charged sites-eliminates EOF, sample adsorption

Dynamic coating of charged sites on fused silica capillary is accomplished with POP-4 polymer

Capillary Wall Coatings Impact DNA Separations



Electroosmotic flow (EOF)

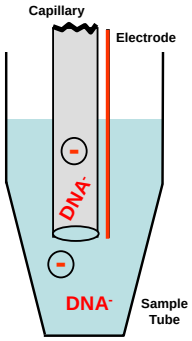
Solvated ions drag solution towards cathode in a flat flow profile

How to Improve Resolution?

1. **Lower Field Strength**
2. **Increase Capillary Length**
3. **Increase Polymer Concentration**
4. **Increase Polymer Length**

All of these come at a cost of longer separation run times

Electrokinetic Injection Process



Amount of DNA injected is inversely proportional to the ionic strength of the solution

Salty samples result in poor injections

Sample Conductivity Impacts Amount Injected

$$[DNA_{inj}] = \frac{Et(\pi r^2)\,(\mu_{ep} + \mu_{eof})[DNA_{sample}]\,(\lambda_{buffer})}{\lambda_{sample}}$$

$[DNA_{inj}]$ is the amount of sample injected

E is the electric field applied

t is the injection time

r is the radius of the capillary

μ_{ep} is the mobility of the sample molecules

μ_{eof} is the electroosmotic mobility

$[DNA_{sample}]$ is the concentration of DNA in the sample

λ_{buffer} is the buffer conductivity

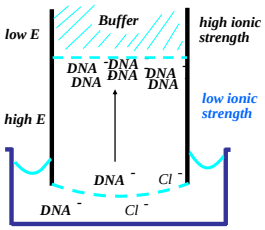
λ_{sample} is the sample conductivity

Cl- ions and other buffer ions present in PCR reaction contribute to the sample conductivity and thus will compete with DNA for injection onto the capillary

Butler *et al.* (2004) Electrophoresis 25: 1397-1412

Two Major Effects of Sample Stacking

1. Sample is preconcentrated. Effect is inversely proportional to ionic strength
2. Sample is focused. Ions stop moving in low electric field
3. Mobility of sample = μ_{ep} = velocity/ electric field



Steps Performed in Standard Module

See J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition; Chapter 14

- **Capillary fill** – polymer solution is forced into the capillary by applying a force to the syringe
- **Pre-electrophoresis –** the separation voltage is raised to 10,000 volts and run for 5 minutes;
- **Water wash of capillary – capillary is dipped several times in deionized water to remove buffer salts that would interfere with the injection process**
- **Sample injection** – the autosampler moves to position A1 (or the next sample in the sample set) and is moved up onto the capillary to perform the injection; a voltage is applied to the sample and a few nanoliters of sample are pulled onto the end of the capillary; the default injection is 15 kV (kilovolts) for 5 seconds
- **Water wash of capillary** – capillary is dipped several times in waste water to remove any contaminating solution adhering to the outside of the capillary
- **Water dip** – capillary is dipped in clean water (position 2) several times
- **Electrophoresis** – autosampler moves to inlet buffer vial (position 1) and separation voltage is applied across the capillary; the injected DNA molecules begin separating through the POP-4 polymer solution
- **Detection** – data collection begins; raw data is collected with no spectral deconvolution of the different dye colors; the matrix is applied during Genescan analysis

Comments on Sample Preparation

- Use high quality formamide (<100 μS/cm)!
  - ABI sells Hi-Di formamide
  - regular formamide can be made more pure with ion exchange resin
- Deionized water vs. formamide
  - Biega and Duceman (1999) *J. Forensic Sci.* 44: 1029-1031
  - *Crivellente, Journal of Capillary Electrophoresis* **2002,** *7 (3-4),* 73-80.
  - water works fine but samples are not stable as long as with formamide; water also evaporates over time…
- Denaturation with heating and snap cooling
  - use a thermal cycler for heating and cold aluminum block for snap cooling
  - heat/cool denaturation step is necessary only if water is substituted for formamide...

January 6, 2005 Letter from Applied Biosystems to ABI 310 Customers

- "Testing has shown that Hi-Di Formamide denatures DNA **without the need to heat samples**…"
- In other words, no heat denaturation and snap cooling needed!

Applied Biosystems Okays Use of Deionized Water for DNA Sequencing

Technical Bulletin #1 **Issued August 2006**

Applied Biosystems 3730/3730*xl* DNA Analyzer

Subject: Influence of Sequencing Injection Solution on 3730/3730xl DNA Analyzer Performance

In this Bulletin:

- Three Loading Solutions Tested on Page 1
- Loading Solution Test Data on Page 2
- Recommendations on Page 6
- Guidelines for Use on Page 6

Three Loading Solutions Tested

Loading Solution Background

Applied Biosystems presently recommends the use of Hi-Di™ Formamide as the sample-loading solution for all Applied Biosystems DNA sequencers to ensure sample preservation and resistance to evaporation. However, many users of the 3730 choose either deionized water or dilute EDTA solutions. These choices are driven largely by cost and safety/hazardous material considerations.

Detection Issues

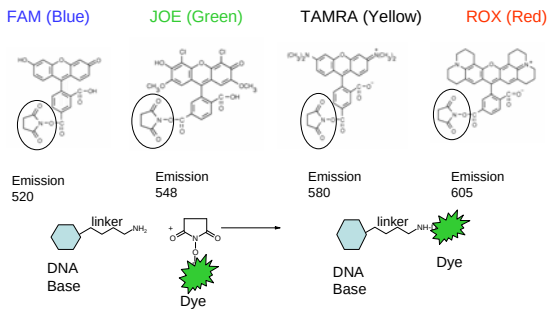
- Fluorescent dyes
  - spectral emission overlap
  - relative levels on primers used to label PCR products
  - dye "blobs" (free dye)
- Virtual filters
  - hardware (CCD camera)
  - software (color matrix)

Filters determine which wavelengths of light are collected onto the CCD camera

Fluorescent Labeling of PCR Products

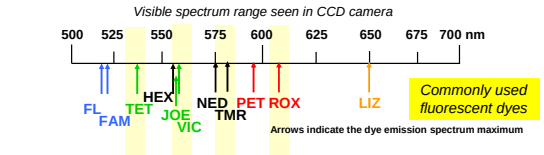
- Dyes are attached to one primer in a pair used to amplify a STR marker
- Dyes are coupled to oligonucleotides (primers) through NHS-esters and amine linkages on the 5'end of the primer: Dye-$(CH_2)_6$-primer
- Dye-labeled oligonucleotides are incorporated during multiplex PCR amplification giving a specific color "tag" to each PCR product
- PCR products are distinguished using CCD imaging on the 310

Amine Reactive Dyes used in Labeling DNA



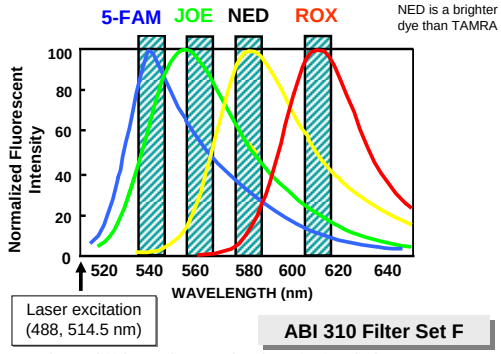
The succinimidyl ester reacts rapidly with amine linkers on DNA bases

Virtual Filters Used in ABI 310



	Blue	Green	Yellow	Red	Orange	Used with These Kits
Filter A	FL	JOE	TMR	CXR		PowerPlex 16
Filter C	6FAM	TET	HEX	ROX		in-house assays
Filter F	5FAM	JOE	NED	ROX		Profiler Plus
Filter G5	6FAM	VIC	NED	PET	LIZ	Identifiler

Fluorescent Emission Spectra for ABI Dyes

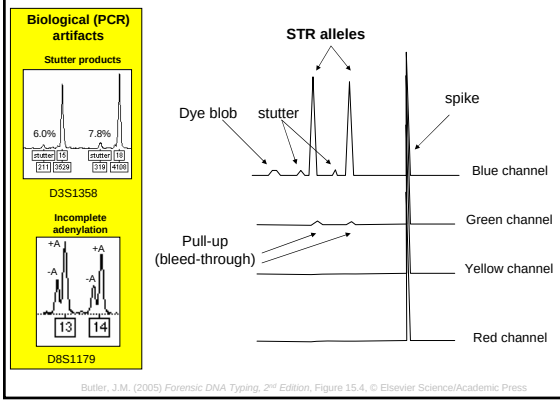


Butler, J.M. (2001) *Forensic DNA Typing*, Figure 10.4, ©Academic Press

Please Note!

- There are no filters in a 310
- Its just the choice of pixels in the CCD detector
- All the light from the grating is collected
- You just turn some pixels on and some off

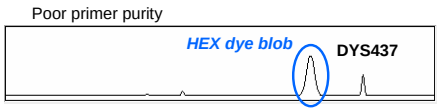
Deciphering Artifacts from the True Alleles



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 15.4, © Elsevier Science/Academic Press

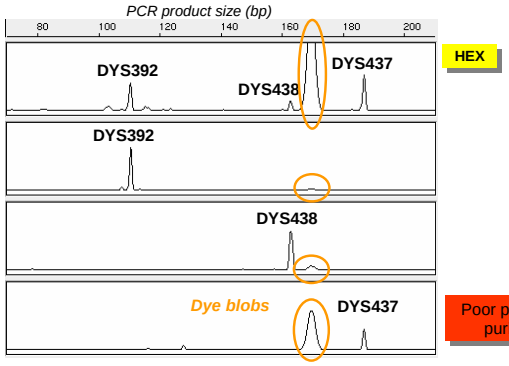
Dye Blobs ("Artifacts")

- Free dye (not coupled to primer) can be injected into the CE capillary and interfere with detection of true STR alleles
- **Dye blobs are wider and usually of less intensity** than true STR alleles (amount depends on the purity of the primers used)
- Dye blobs usually appear at an apparent size that is unique for each dye (e.g., FAM ~120 bp, PET ~100 bp)

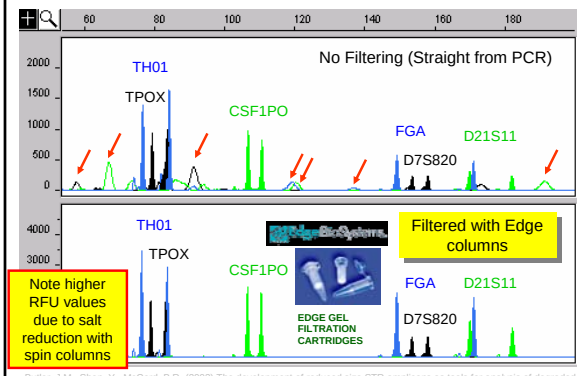


Dye Blob Problems with Some PCR Primers

Individual Y-STR Locus Amplifications



Removal of Dye Artifacts Following PCR Amplification



Butler, J.M., Shen, Y., McCord, B.R. (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci* 48(5) 1054-1064.

Conclusions

DNA typing by capillary electrophoresis involves:

1) The use of entangled polymer buffers
2) Injection by sample stacking
3) Multichannel laser induced fluorescence
4) Internal and external calibration

Practical Aspects of ABI 310/3100 Use

ABI Genetic Analyzer Usage at NIST

- ABI 310 x 2 (originally with Mac, then NT)
  - 1st was purchased in 1996
  - 2nd was purchased in June 2002
- ABI 3100 (Data collection v1.0.1)
  - Purchased in June 2002
  - Original data collection software retained
- ABI 3130xl upgrade (Data collection v3.0)
  - Purchased in April 2001 as ABI 3100
  - Upgraded to ABI 3130xl in September 2005
  - Located in a different room

Jan 2007 – upgraded to 3130xl with data collection v3.0

Our Use of the ABI 3100

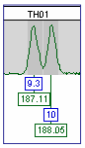
- Data collection software, version 1.0.1
- **POP-6** with 36 cm capillary array

- STR kits and in-house assays for autosomal STRs, Y-STRs, and miniSTRs
- SNaPshot assays for mtDNA SNPs, Y-SNPs, and autosomal SNPs
- DNA sequencing for mtDNA and STR repeat sequencing

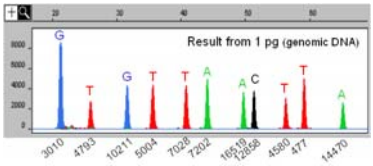
We can routinely get more than 400 runs per capillary array by not changing the polymer between applications

NIST ABI 3100 Analysis Using POP-6 Polymer

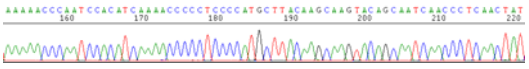
High Resolution STR Typing



SNaPshot SNP Typing
(Coding Region mtSNP 11plex minisequencing assay)

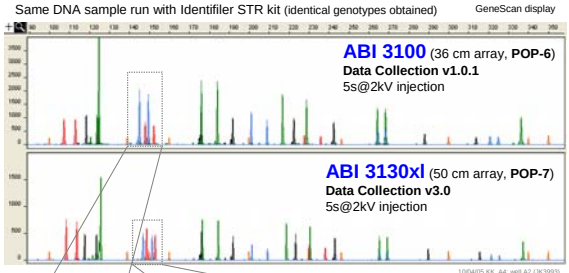


mtDNA Sequencing (HV1)



Comparison of ABI 3100 Data Collection Versions

Same DNA sample run with Identifiler STR kit (identical genotypes obtained) GeneScan display



v1.0.1 v3.0

Relative peak height differences are due to "variable binning" with newer ABI data collection versions.

Difference in the STR allele relative mobilities (peak positions) are from using POP-6 vs. POP-7.

Consumables for ABI 310/3100

What we use at NIST

- A.C.E.™ Sequencing Buffer 10X (Amresco)
  - $155/L = $0.0155/mL 1X buffer **(costs 20 times less!)**
  - http://www.amresco-inc.com
- 3700 **POP-6 Polymer** (Applied Biosystems)
  - $530 / 200 mL = $2.65/mL **(costs 20 times less!)**

What ABI protocols suggest

- 10X Genetic Analyzer Buffer with EDTA
  - $78/25 mL = $0.312/mL 1X buffer (ABI)
- 3100 POP-4 Polymer
  - $365 / 7 mL = $52/mL

2004 prices

Maintenance of ABI 310/3100/3130

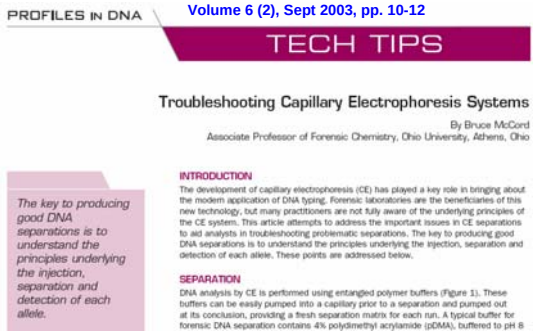
- Syringe – leaks cause capillary to not fill properly
- Capillary storage & wash – **it dries, it dies!**
- Pump block – cleaning helps insure good fill
- Change the running buffer regularly

YOU MUST BE CLEAN AROUND A CE!

Overall Thoughts on the ABI 310/3100/3130

- Settling on a common instrument platform has been good for the forensic DNA community in terms of data consistency (this is also true with the use of common STR kits)
- I am concerned that the community is very dependent primarily on one company…
- I really like using the instrument and can usually get nice data from it
- Like any instrument, it has its quirks…

Bruce McCord's *Profiles in DNA* Article



PROFILES IN DNA

Volume 6 (2), Sept 2003, pp. 10-12

TECH TIPS

Troubleshooting Capillary Electrophoresis Systems

By Bruce McCord
Associate Professor of Forensic Chemistry, Ohio University, Athens, Ohio

The key to producing good DNA separations is to understand the principles underlying the injection, separation and detection of each allele.

INTRODUCTION

The development of capillary electrophoresis (CE) has played a key role in bringing about the modern application of DNA typing. Forensic laboratories are the beneficiaries of this new technology, but many practitioners are not fully aware of the underlying principles of the CE system. This article attempts to address the important issues in CE separations to aid analysts in troubleshooting problematic separations. The key to producing good DNA separations is to understand the principles underlying the injection, separation and detection of each allele. These points are addressed below.

SEPARATION

DNA analysis by CE is performed using entangled polymer buffers (Figure 1). These buffers can be easily pumped into a capillary prior to a separation and pumped out at its conclusion, providing a fresh separation matrix for each run. A typical buffer for forensic DNA separation contains 4% polydimethyl acrylamide (pDMA), buffered to pH 8

CE Troubleshooting

Bruce McCord, AAFS 2006 Workshop (Seattle, WA)
February 20, 2006

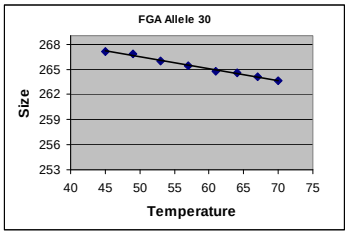
Outline for This Section

1. Chemistry/molecular biology problems – stutter, -A, degradation, inhibition, low copy #
2. Sample and buffer problems – formamide, urea, water, salt concentration, free dye ("dye blobs")
3. External factors – power supply, room temperature, cleanliness, voltage leaks
4. Instrument problems – optical system, capillary clogging, air bubbles, syringe leaks
5. Troubleshooting benchmarks/QC monitoring

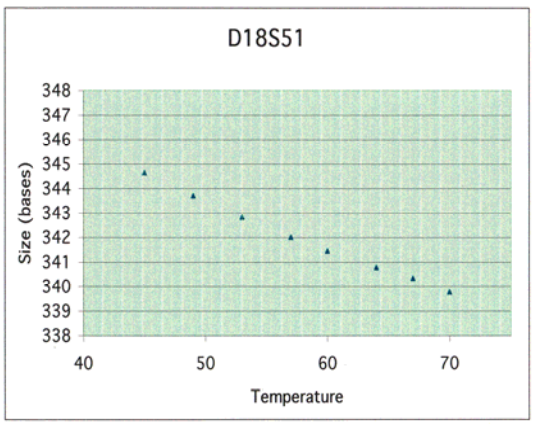
3. External Factors

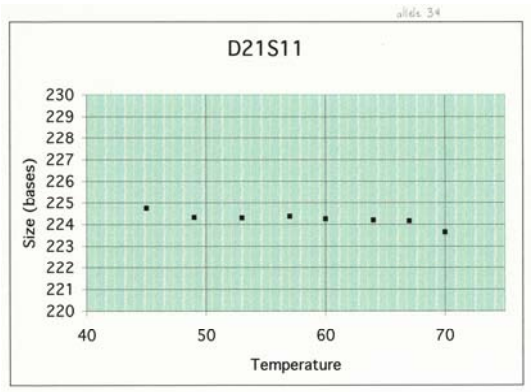
- Room temperature
  - Variations in room temperature can cause mobility shifts with band shifts and loss of calibration
  - Temperature is also important due to effects of high humidity on electrical conductance
- Cleanliness
  - Urea left in sample block can crystallize and catalyze further crystal formation causing spikes, clogs and other problems.
  - Best bet is to keep polymer in system and not remove or change block until polymer is used up.

Effect of Temperature on allele size

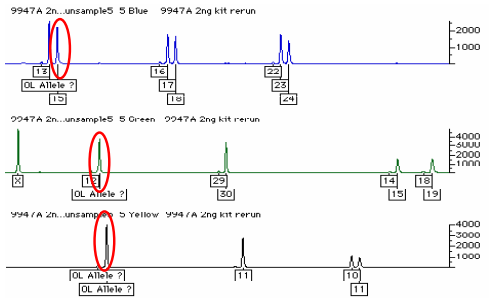


Slope is 0.14 bases/degree centigrade
Therefore a small change in temperature has a big effect
(A 1-2 degree shift in temperature of the heat plate can produce an OL allele)

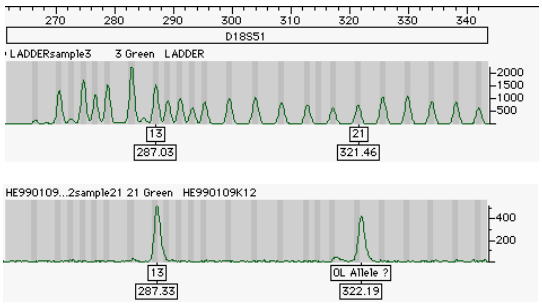




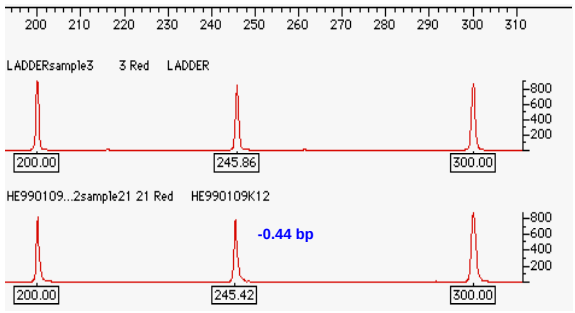
Temperature Effects: "OL" Alleles



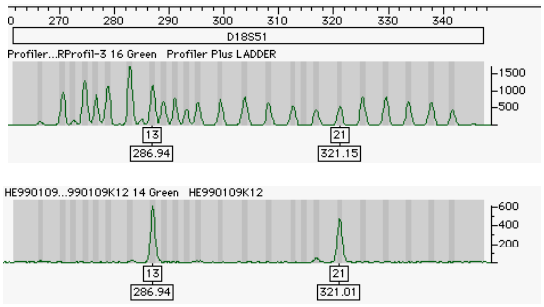
"OL Alleles"



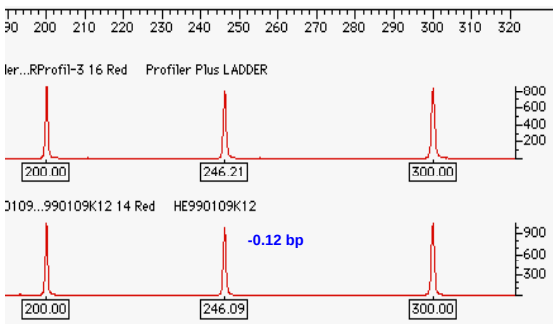
"OL alleles " - look at the 250 peak



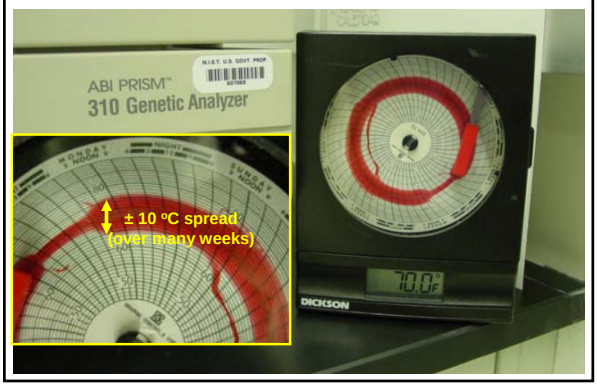
"OL allele re-injected"



And the 250 peak...



Monitoring Room Temperature Over Time



Temperature Probes

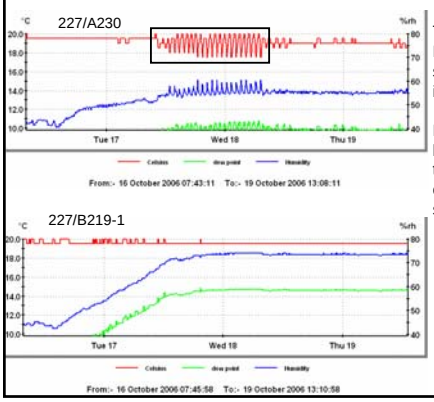
Refrigerator and freezer monitoring

Frig/Freeze Monitors $240
#DT-23-33-80 – USB Temperature Datalogger PLUS Software $79.00 (#DT-23-33-60)

Room Monitors, # DT-23039-52 – USB Temperature-Humidity Datalogger $91.00
(Cole Parmer, Vernon Hills IL)

Room temperature monitoring

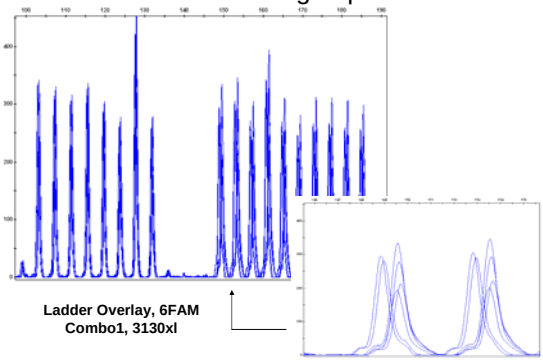
Monitoring Instrument Room Temperature Fluctuations



Temperature Monitoring of two separate instrument rooms.

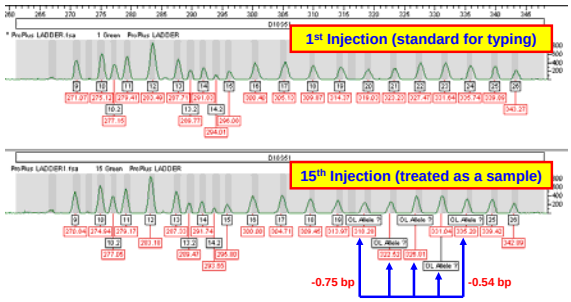
Box area is a 24 hour period where temperature control is not stable.

Poor Temperature Control Causes DNA Sizing Imprecision



Ladder Overlay, 6FAM Combo1, 3130xl

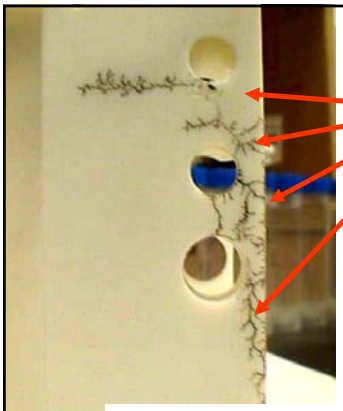
Use of Second Allelic Ladder to Monitor Potential Match Criteria Problems



These alleles have drifted outside of their genotyping bins due to temperature shifting over the course of the sample batch

Cleanliness

- Urea sublimates and breaks down to ionic components - these find a path to ground
- Similarly wet buffer under a vial creates paths to ground
- Capillary windows must be clear or matrix effects will occur
- Laser will often assist in this process
- Vial caps will transfer low levels of DNA to capillary



Carbon Trails

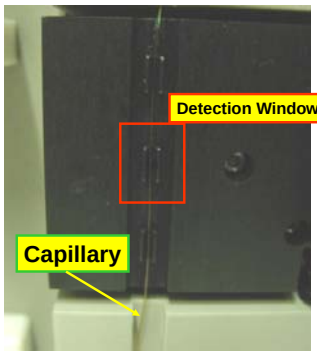
High Humidity or wet buffer vials can create other paths to ground

Keep Your System Clean!

4. Instrumental Factors

- Optical System
  - Sensitivity changes with age, capillary diameter, capillary cleanliness, instrument calibration
- Fluidic System
  - Effects of bubbles, dust, urea crystals, leaks in syringe and capillary ferrule
- Matrix Calculations
  - Changes in buffer, optics, sample dye can alter the software calibrations
- Capillary Problems
  - Chemisorbed materials on capillary surface can produce osmotic flow, DNA band broadening and inconsistent resolution (meltdowns)

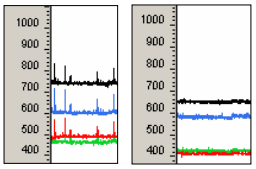
The Detection Window



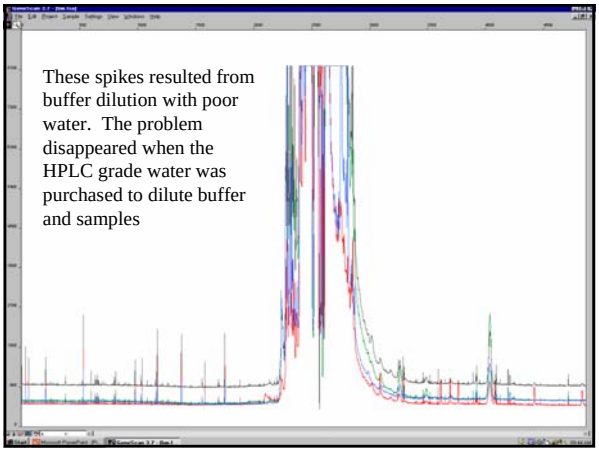
Make sure that the capillary window is lined up (if it is not, then no peaks will be seen)

Window may need to be cleaned with ethanol or methanol

Review Start of Raw Data Collection

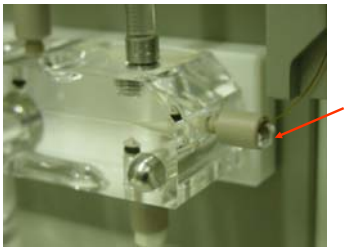


Little spikes indicate need to change buffer… check current



These spikes resulted from buffer dilution with poor water. The problem disappeared when the HPLC grade water was purchased to dilute buffer and samples

Beware of Urea Crystals



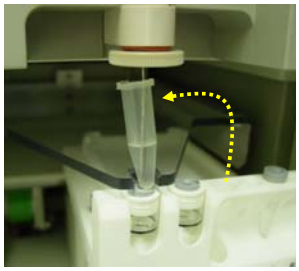
Urea crystals have formed due to a small leak where the capillary comes into the pump block

Urea sublimates and can evaporate to appear elsewhere

Use a small balloon to better grip the ferrule and keep it tight

Pump block should be well cleaned to avoid problems with urea crystal formation

Storage when ABI 310 is not in use



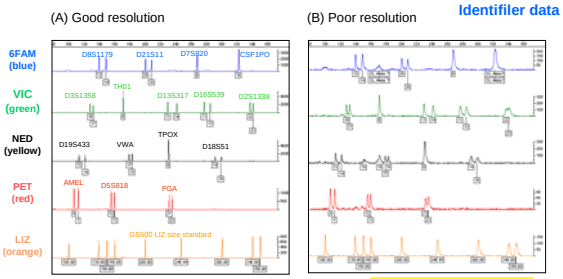
Remember that the water in the open tube will evaporate over time…

- Keep inlet of capillary in water…if it dries out then urea crystals from the polymer will clog the opening
- The waste vial (normally in position 3) can be moved into position
- A special device can be purchased from Suppelco to rinse the capillary off-line
- Store in distilled water
- Note that the laser is on when the instrument is on

Buffer Issues

- The buffer and polymer affect the background fluorescence- affecting the matrix
- Urea crystals and dust may produce spikes
- High salt concentrations may produce reannealing of DNA
- High salt concentrations affect current
- Low polymer concentrations affect peak resolution

Capillary Meltdowns



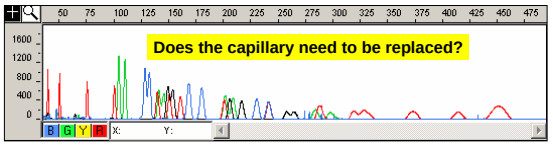
Good Capillary in 3100 Array

Bad Capillary in 3100 Array

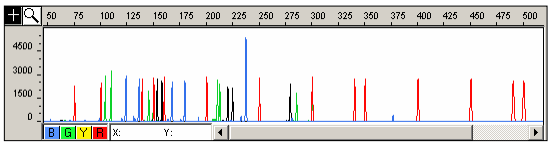
Butler, J.M., Buel, E., Crivellente, F., McCord, B.R. (2004) Forensic DNA typing by capillary electrophoresis: using the ABI Prism 310 and 3100 Genetic Analyzers for STR analysis. *Electrophoresis*, 25: 1397-1412.

Meltdowns can be permanent or transitory

as we have seen these may result from sample contamination effects



No! The next injection looks fine…



Meltdowns may be the result of

- Bad formamide
- Excess salt in sample/renaturation
- Water in the polymer buffer
- **Syringe leak** or bottom out
- Poisoned capillary
- Conductive polymer buffer due to urea degradation
- Crack/shift in capillary window
- Detergents and metal ions

5. Troubleshooting benchmarks

- **Monitor run current**
- Observe syringe position and movement during a batch
- Examine ILS (ROX) peak height with no sample
- Observe “250 bp” peak in GS500 size standard
- Monitor resolution of TH01 9.3/10 in allelic ladder and size standard peak shapes
- **Keep an eye on the baseline signal/noise**
- Measure formamide conductivity
- Reagent blank – are any dye blobs present?
- See if positive control DNA is producing typical peak heights (along with the correct genotype)

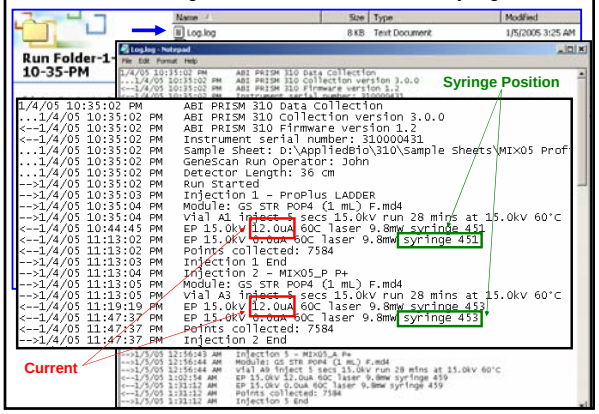
Measurement of Current

- V/I = R where R is a function of capillary diameter, [buffer], and buffer viscosity
- In a CE system the voltage is fixed, thus changes in resistance in the capillary will be reflected in the current observed
- Air bubbles, syringe leaks, alternate paths to ground, changes in temperature, changes in zeta potential, and contamination, will be reflected in the current
- A typical current for a CE system with POP4 buffer is **8-12 µA** (microamps)

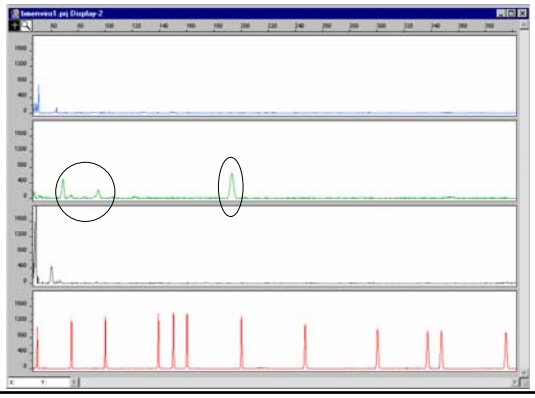
Syringe Travel

- The ABI 310 instrument also keeps track of the position of the syringe (in the log file)
- Depending on the resistance to flow, the syringe will travel different lengths
- Syringe leaks may be reflected in a longer distance traveled prior to each injection
- These leaks occur around the barrel of the syringe and at the connection to the capillary block

Use of ABI 310 Log File to Monitor Current and Syringe Travel



Dye Blobs in the Negative Control Sample



Measuring Formamide Conductivity

(not this way)

The key is to measure the bottle when it comes in or buy the good stuff and immediately pipette it out into small tubes with or without ROX already added. Then freeze the tubes.

Do not ever open a cold bottle of formamide. Water will condense inside and aid in the formation of conductive formic acid.

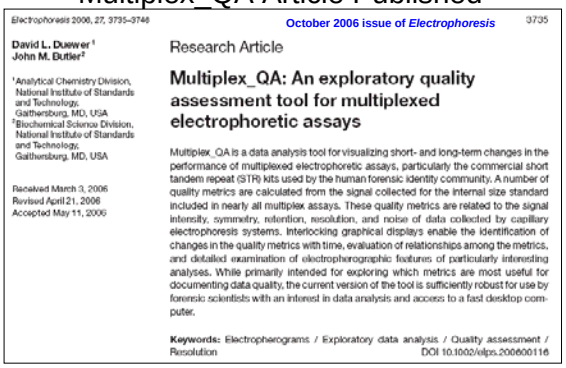
Conclusion:

Troubleshooting is more than following the protocols

It means keeping watch on all aspects of the operation

1. Monitoring conductivity of sample and formamide
2. Keeping track of current and syringe position in log.
3. Watching the laser current
4. Watching and listening for voltage spikes
5. Monitoring room temperature and humidity

Multiplex_QA Article Published



Electrophoresis 2006, 27, 3735–3746 — October 2006 issue of *Electrophoresis* — 3735

David L. Duewer[1] John M. Butler[2]

[1]Analytical Chemistry Division, National Institute of Standards and Technology, Gaithersburg, MD, USA
[2]Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

Received March 3, 2006
Revised April 21, 2006
Accepted May 11, 2006

Research Article

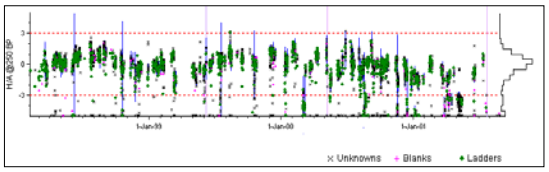
Multiplex_QA: An exploratory quality assessment tool for multiplexed electrophoretic assays

Multiplex_QA is a data analysis tool for visualizing short- and long-term changes in the performance of multiplexed electrophoretic assays, particularly the commercial short tandem repeat (STR) kits used by the human forensic identity community. A number of quality metrics are calculated from the signal collected for the internal size standard included in nearly all multiplex assays. These quality metrics are related to the signal intensity, symmetry, retention, resolution, and noise of data collected by capillary electrophoresis systems. Interlocking graphical displays enable the identification of changes in the quality metrics with time, evaluation of relationships among the metrics, and detailed examination of electropherographic features of particularly interesting analyses. While primarily intended for exploring which metrics are most useful for documenting data quality, the current version of the tool is sufficiently robust for use by forensic scientists with an interest in data analysis and access to a fast desktop computer.

Keywords: Electropherograms / Exploratory data analysis / Quality assessment / Resolution — DOI 10.1002/elps.200600116

User manual (127 pages) available for download from STRBase

Multiplex_QA Overview



- **Research tool** that **provides quality metrics to review instrument performance over time** (e.g., examines resolution and sensitivity using internal size standard peaks)
- Runs with Microsoft Excel macros. Requires STR data to be converted with NCBI's BatchExtract program into numerical form.

Available for download from STRBase:
http://www.cstl.nist.gov/biotech/strbase/software.htm

Acknowledgments

NIST Human Identity Project Team

Leading the Way in Forensic DNA…

John Butler (Leader) | Margaret Kline | Pete Vallone | Jan Redman | Amy Decker | Becky Hill | Dave Duewer

Funding from interagency agreement 2003-IJ-R-029 between the National Institute of Justice and the NIST Office of Law Enforcement Standards

Many wonderful collaborators from industry, university, and government laboratories.

Bruce McCord (Florida International University) for many of the slides

Thank you for your attention…

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

Questions?

See also http://www.dna.gov/research/nist

http://www.cstl.nist.gov/biotech/strbase
john.butler@nist.gov

Mixture Interpretation Questions

Homework for Monday Night

Name (*leave blank if you want to be anonymous*): _____ Email address:_____

Interpretation Guidelines

What would you like to see in national guidelines on how to perform DNA mixture interpretation and statistical analysis?

How does your lab handle reference samples during interpretation of evidence? Do you try to solve the mixture entirely without looking at either victim or suspect profiles?

What kind of pre-case assessment do you perform when approaching a case where a possible mixture is involved?

Does your lab attempt statistics on a minor component? If so, what types of statistics are used?

Do you have a decision point whereby you consider a mixture too complicated and do not try to solve it? How do you know when to stop in terms of mixture interpretation?

Are composite profiles acceptable – e.g., high injection for minor component and low injection for major component allele identification?

How do you report mixture statistics in court?

Would a flowchart for mixture interpretation be helpful?

Validation and Training

For your lab validation studies of a new STR kit or instrument, how many mixtures should be evaluated? How do you decide on what combination of alleles to include in such a study?

What kind of training materials would be beneficial to help your laboratory more effectively solve mixtures?

Suggestions for training staff to have more analyst consistency within your lab:

Other Topics

What percentage of time is spent in a case trying to deduce the mixture components?

Have you seen performance differences between various STR typing kits that would impact mixture interpretation?

Is your lab using Y-STRs to help with mixtures?

What kinds of software features would be valuable to aid mixture interpretation?

What are the biggest obstacles you face in your lab in terms of mixture interpretation?

Mixture Interpretation Discussion

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

NIST and NIJ Disclaimer

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

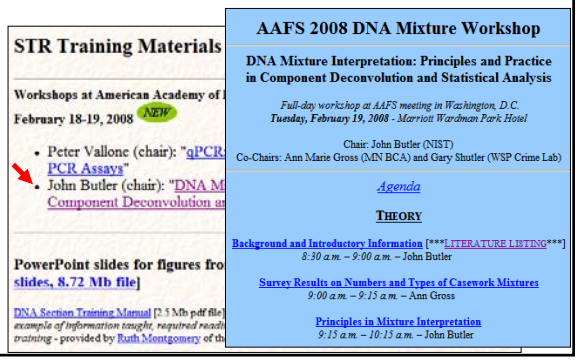
Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

SWGDAM Disclaimer…

Training Information Available on STRBase

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



AAFS 2008 Workshop Presenters

Ann Marie Gross
MN BCA

John M. Butler
NIST

George Carmody
Carleton University/ Statistical Consultant

Gary Shutler
Wash State Police Crime Lab

Angie Dolph
Marshall University (NIST Summer Intern)

Joanne B. Sgueglia
Mass State Police Crime Lab

Tim Kalafut
US Army Crime Lab

Purpose for Teaching AAFS Workshop

We hope that participants:

- Gain a better understanding of the current approaches being used throughout the community for mixture interpretation
- See worked examples of mixture component deconvolution and statistical analysis
- Come away with ideas to improve your laboratory's interpretation guidelines and training regarding mixtures in forensic casework

AAFS Workshop Morning Agenda - Theory

Background and Introductory Information
8:30 a.m. – 9:00 a.m. – John Butler

Survey Results on Numbers and Types of Casework Mixtures
9:00 a.m. – 9:15 a.m. – Ann Gross

Principles in Mixture Interpretation
9:15 a.m. – 10:15 a.m. – John Butler

***10:15 a.m. – 10:30 a.m.* BREAK**

Strategies for Mixture Deconvolution with Worked Examples
10:30 a.m. – 11:30 a.m. – John Butler

Different Approaches to Statistical Analysis of Mixtures
11:30 a.m. – 12:00 p.m. – George Carmody

***12:00 p.m. – 1:15 p.m.* LUNCH**

Afternoon Agenda – **Practical Application**

Real Case Example – Importance of Properly Stating Your Conclusions
1:15 p.m. – 1:30 p.m. – Gary Shutler

Variability between Labs in Approaches & Mixture Interlaboratory Studies
1:30 p.m. – 2:15 p.m. – John Butler

Validation Studies and Preparing Mixture Interpretation Guidelines
2:15 p.m. – 2:45 p.m. – Joanne Sgueglia

***2:45 p.m. – 3:00 p.m.* BREAK**

Testing of Mixture Software Programs
3:00 p.m. – 3:15 p.m. – Angela Dolph

DNA_DataAnalysis Software Demonstration
3:15 p.m. – 4:00 p.m. – Tim Kalafut

Training Your Staff to Consistently Interpret Mixtures
4:00 p.m. – 4:45 p.m. – Panel Discussion with Ann Gross, Gary Shutler, Joanne Sgueglia

***4:45 p.m. – 5:00 p.m.* – Questions and Answers as needed**

Recent Mixture Workshops
Conducted by John Butler

Helpful feedback obtained from workshop participants

Southern Association of Forensic Scientists (SAFS)
September 11, 2007 (Atlanta, GA)
- **Mixture Interpretation (theory)**
- Along with Software discussion (Rhonda Roby) and demonstration (Tom Overson/Tim Kalafut)
- **33 attendees from 13 different labs**

Northeastern Association of Forensic Scientists (NEAFS)
November 2-3, 2007 (Bolton Landing, NY)
- **The Cutting Edge of DNA Testing: Mixture Interpretation, miniSTRs, and Low Level DNA**
- **42 attendees from 13 different labs**

NEAFS Workshop materials (70 pages) available on STRBase:
http://www.cstl.nist.gov/biotech/strbase/pub_pres/NEAFS2007_CuttingEdgeDNA.pdf

Mixture Basics

From J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*, p. 154

- Mixtures arise when two or more individuals contribute to the sample being tested.
- Mixtures can be challenging to detect and interpret without extensive experience and careful training. **Even more challenging with poor quality data when degraded DNA is present…**
- Differential extraction can help distinguish male and female components of many sexual assault mixtures. **Y-chromosome markers can help here in some cases…**

More on Mixtures...

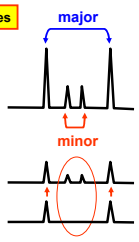
Most mixtures encountered in casework are 2-component mixtures arising from a combination of victim and perpetrator DNA profiles

Ann Gross will discuss some recent collected casework summaries

Torres *et al.* (2003) *Forensic Sci. Int.* 134:180-186 examined 1,547 cases from 1997-2000 containing **2,424 typed samples** of which **163 (6.7%) contained a mixed profile** with **only 8 (0.3%) coming from more than two contributors**

95.1% (155/163) were 2-component mixtures

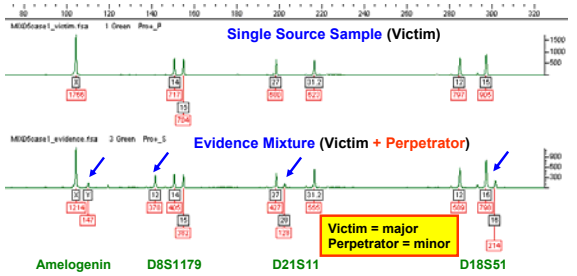
Ratios of the various mixture components stay fairly constant between multiple loci enabling deduction of the profiles for the major and minor components



Some mixture interpretation strategies involve using victim (or other reference) alleles to help isolate obligate alleles coming from the unknown portion of the mixture

http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm
MIX05 Case #1; Profiler Plus green loci

Example Mixture Data (MIX05 Study-Profiler Plus)



Obligate Alleles (not present in the victim reference)

Y	12	28	16

True "Perpetrator" Profile

X,Y	12,12	28,31.2	15,16

Sources of DNA Mixtures

- **Two (or more) individuals** contribute to the biological evidence examined in a forensic case (e.g., sexual assault with victim and perpetrator or victim, consensual sexual partner, and perp)

  Victim Reference and Spouse or Boyfriend Reference
- **Contamination** of a single source sample from
  - evidence collection staff
  - laboratory staff handling the sample
  - Low-level DNA in reagents or PCR tubes or pipet tips

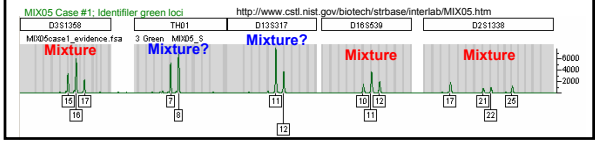
  Examine Staff Profiles (Elimination Database), etc.

Reference elimination samples are useful in deciphering both situations due to possibility of intimate sample profile subtraction

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.

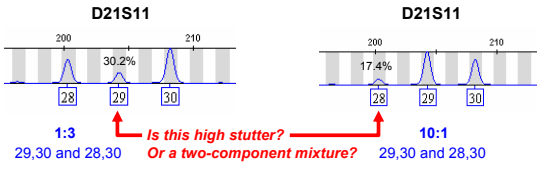


Detecting Mixtures

- Review and compile information from the entire profile – **don't just focus on a single locus!**
- **Tri-allelic patterns exist** in single source samples
  - **145 different tri-alleles recorded for the 13 core CODIS loci** on STRBase as of Jan 22, 2008
  - CSF1PO (5), FGA (22), TH01 (1), TPOX (15), VWA (18), D3S1358 (6), D5S818 (4), D7S820 (7), D8S1179 (11), D13S317 (8), D16S539 (8), D18S51 (21), D21S11 (19)
- A mixture often declared when **>2 peaks in ≥2 loci**

Mixtures: Issues and Challenges

- Artifacts of PCR amplification such as stutter products and heterozygote peak imbalance complicate mixture interpretation
- Thus, only a limited range of mixture component ratios can be solved routinely



1

Responses to Questions
from a Previous Mixture Workshop (Fall 2007)

What are the biggest obstacles you face in your lab in terms of mixture interpretation?

- Trying to be consistent in my interpretation and with coworkers
- Consistency between analysts
- No consistency – based on analysts discretion/experience; due to lack of consistent training
- Vague SOP leading to inconsistency between analysts due to differences in how "conservative" or not each analyst is
- There is a lot of "individual interpretation" in our lab
- Varying opinions between interpreting analysts due to lack of uniform guidelines
- Resistance to change from other analysts/supervisors
- Getting management to commit to guidelines that will be followed by everyone

2

Responses to Questions
from a Previous Mixture Workshop (Fall 2007)

What are the biggest obstacles you face in your lab in terms of mixture interpretation?

- Where to draw the line without throwing away valuable data
- Partial minor contributors
- Stochastic effects in minor components
- STATS and presenting them in court so that the jury will understand them
- When to do stats and what stats to do in different cases
- Lack of concrete/uniform guidelines from statisticians

DNA Mixture Interpretation:
Principles and Practice in Component Deconvolution and Statistical Analysis

Numbers and Types of Casework Mixtures

Handouts available on STRBase at
http://www.cstl.nist.gov/biotech/strbase/training/AAFS2008_MixtureWorkshop.htm

AAFS 2008 Workshop #16
Washington, DC
February 19, 2008

Ann Marie Gross
ann.gross@state.mn.us

Mixtures……

- How often are mixtures obtained
- What types of mixtures are we seeing
  - Where should we focus our attention for training
  - What info can we give to the forensic community regarding mixtures
- What types of samples most often yield mixtures

Torres et al. 4 year Spanish study

- Four year study (1/1997 to 12/2000)
- 2412 samples typed
  - 955 samples from sexual assaults
  - 1408 samples from other offenses
  - 49 samples from human remains identifications
- 163/2412 samples (6.7% showed mixed profile)

Spreadsheet Information Requested

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

Labs requested to also provide info on kit, PCR volume used, etc.

- **Case#**
- **Item#**

This information retained by lab and not returned…

- **Type of sample (biological material if ID'd)**
- **Type of substrate**
- **Quantity amp'd**
- **Minimum # of contributors (1, 2, 3, 4, or >4)**
- **Predominant type (major profile) determined?**
- **Stats reported**
- **Comments**

We would love to have your lab mixture numbers…

Email information to Ann.Gross@state.mn.us

12 Labs Submitted Data
(prior to AAFS meeting)

- **Palm Beach Sheriff's Office Crime Lab, Florida**
- Centre for Forensic Science, Toronto
- Connecticut State Police
- Washington State Police
- New Jersey State Police
- Georgia Bureau of Investigation
- Royal Canadian Mounted Police, Ottawa
- USACIL, Georgia
- Michigan State Police
- Kern County Crime Lab, California
- CAL DOJ
- Minnesota Bureau of Criminal Apprehension

We would still like to collect more case summary data…

All Laboratory Data Combined

Case type	N = 3106	1	2	3	4	>4
			# contributors			
Sexual Assault	N = 1408	51%	40%	8%	--	--
Major Crime	N = 1388	66%	24%	8%	2%	--
High Volume	N = 310	43%	37%	19%	1%	--
		Single source	Mixtures			

Overall Summary – 3106 samples

- 57% of samples from all types of cases are single source
- 43% of samples from all types of cases are mixtures
  - 33% of mixtures of at least two contributors
  - 9% of mixtures of at least three contributors
  - 1% of mixtures of at least four contributors

Focus in training materials will be on two-person mixtures as they presently predominate

Principles of Mixture Interpretation

Topics for Discussion

- SWGDAM Mixture Interpretation Committee progress
- Different statistical approaches: CPE or LR
- ISFG Mixture Interpretation Recommendations
  - UK response
  - German categories for mixtures
- Validation as it relates to mixture interpretation
  - Stochastic threshold vs analytical threshold
- Low-level DNA and mixtures
- Important elements of interpretation guidelines

SWGDAM Mixture Interpretation Subcommittee

- **John Butler** (NIST) - chair
- **Gary Sims** (CA DOJ) - co-chair
- Mike Adamowicz (CT)
- Jack Ballantyne (UCF/NCFS)
- George Carmody (Carleton U)
- Cecelia Crouse (PBSO)
- Allison Eastman (NYSP)
- Roger Frappier (CFS-Toronto)
- Ann Gross (MN BCA)
- Phil Kinsey (MT)
- Jeff Modler (RCMP)
- Gary Shutler (WSP)

Everyone not at every meeting…

Have met 3 times:
Jan 2007
July 2007
Jan 2008

Through the Jan 2008 meeting we have also had to deal with Y-STR issues – which has limited our focus on mixtures

Additional Participants (Jan 2008)
Bruce Heidebrecht (MD)
Steve Lambert (SC)

Started in January 2007

Progress and Plans for Mixture Committee

- **Guidelines in process** of being discussed and written
- Collecting data on number and type of mixture cases observed in various labs
- Plan to create a training workbook with worked examples
- Considering flow charts to aid mixture interpretation
- Have discussed responses to ISFG Recommendations

I invite your input as to what should be included in the guidelines…

Your HOMEWORK…

Elements of DNA Mixture Interpretation

Principles (theory)

↓

Protocols (validation)

↓

Practice (training & experience)

Consistency across analysts

ISFG Recommendations
SWGDAM Guidelines

⋮

Your Laboratory SOPs

↕

Training within Your Laboratory

We discussed and would advocate periodic training to aid accuracy and efficiency within your laboratory.

ISFG Recommendations on Mixture Interpretation

July 13, 2006 issue of *Forensic Science International*

Our discussions have highlighted a significant need for continuing education and research into this area.

ELSEVIER

Forensic Science International 160 (2006) 90–101

www.elsevier.com/locate/forsciint

DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

P. Gill [a,*], C.H. Brenner [b], J.S. Buckleton [c], A. Carracedo [d], M. Krawczak [e], W.R. Mayr [f], N. Morling [g], M. Prinz [h], P.M. Schneider [i], B.S. Weir [j]

[a] Forensic Science Service, Trident Court, 2960 Solihull Parkway, Birmingham, UK
[b] Forensic Science Group, School of Public Health, University of California, Berkeley, CA 510-330-1911, USA

Abstract

The DNA commission of the International Society of Forensic Genetics (ISFG) was convened at the 21st congress of the International Society for Forensic Genetics held between 13 and 17 September in the Azores, Portugal. The purpose of the group was to agree on guidelines to encourage best practice that can be universally applied to assist with mixture interpretation. In addition the commission was tasked to provide guidance on low copy number (LCN) reporting. Our discussions have highlighted a significant need for continuing education and research into this area. We have attempted to present a consensus from experts but to be practical we do not claim to have conveyed a clear vision in every respect in this difficult subject. For this reason, we propose to allow a period of time for feedback and reflection by the scientific community. Then the DNA commission will meet again to consider further recommendations.

Keywords: STR typing; Biostatistical analysis; Likelihood ratio; Probability of exclusion; Mixtures; ISFG DNA commission

Who is the ISFG
and why do their recommendations matter?

International Society of Forensic Genetics
http://www.isfg.org/

- An international organization responsible for the promotion of scientific knowledge in the field of genetic markers analyzed with forensic purposes.
- Founded in 1968 and represents more than 1100 members from over 60 countries.
- **A DNA Commission regularly offers recommendations on forensic genetic analysis.**

DNA Commission of the ISFG

- DNA polymorphisms (1989)
- PCR based polymorphisms (1992)
- Naming variant alleles (1994)
- Repeat nomenclature (1997)
- Mitochondrial DNA (2000)
- Y-STR use in forensic analysis (2001)
- Additional Y-STRs - nomenclature (2006)
- **Mixture Interpretation (2006)**
- Disaster Victim Identification (2007)

http://www.isfg.org/Publications/DNA+Commission

ISFG Executive Committee

President
Niels Morling
(Copenhagen, Denmark)

Vice-President
Peter Schneider
(Köln, Germany)

Working Party Representative
Mecki Prinz
(New York City, USA)

Treasurer
Leonor Gusmão
(Porto, Portugal)

Secretary
Wolfgang Mayr
(Vienna, Austria)

Angel Carracedo
***FSI Genetics* Editor-in-Chief**
(former ISFG President, VP)
(Santiago de Compostela, Spain)

Authors of ISFG Mixture Article

Peter Gill
Pioneer of forensic DNA techniques and applications
UK's Forensic Science Service (1978-2008)
University of Strathclyde (Apr 2008 – present)

The Statisticians

Charles Brenner
DNA-View,
Berkeley, CA, USA

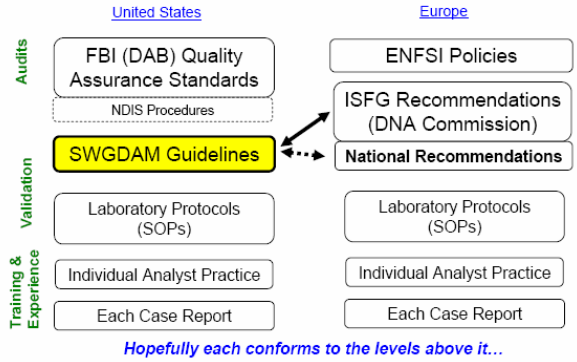
John Buckleton
ESR,
Auckland, New Zealand

Michael Krawczak
Christian-Albrechts-University,
Kiel, Germany

Bruce Weir
U. Washington,
Seattle, USA

My perspective…

Hierarchy of Rules for Forensic DNA Labs



Hopefully each conforms to the levels above it…

UK Response to ISFG Mixture Recommendations

Gill, P., et al. (2008) National recommendations of the technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *FSI Genetics* 2(1): 76–82

Using the published UK response as a model, let us review the nine ISFG Recommendations on mixture interpretation…

From Report to the **Virginia Scientific Advisory Committee** by the DNA Subcommittee – Addendum January 8, 2008

(authored by Dr. Norah Rudin and Dr. Artie Eisenberg)

- "Among the many reasons that Forensic DNA analysis has become the gold standard for forensic science is the relatively discrete nature of the data. For strong, single source samples, a profile can readily be determined, and is subject to little or no analyst judgment. **However, ambiguity may arise when interpreting more complex samples, such as those containing multiple contributors, of poor quality (e.g. degraded or inhibited DNA), of low quantity (e.g. contact samples), or various combinations of these challenging situations**…"

http://www.dfs.virginia.gov/about/minutes/saCommittee/20080108.pdf

From Report to the **Virginia Scientific Advisory Committee** by the DNA Subcommittee – Addendum January 8, 2008

(authored by Dr. Norah Rudin and Dr. Artie Eisenberg)

- "…**These kinds of samples are encountered with increasing frequency, as the sensitivity of the technology has increased, and as law enforcement has become more sophisticated about the kinds of samples they submit for analysis**. Difficult samples are also frequently encountered when reanalyzing historical cases, in which samples were not collected and preserved using the precautions necessary for DNA analysis…"

"Cold cases" or Innocence Project samples…

http://www.dfs.virginia.gov/about/minutes/saCommittee/20080108.pdf

From Report to the **Virginia Scientific Advisory Committee** by the DNA Subcommittee – Addendum January 8, 2008

(authored by Dr. Norah Rudin and Dr. Artie Eisenberg)

- "It is for these types of challenging samples, where the evidence profile may not exactly "match" a reference profile, that confirmation bias becomes a concern. **The interpretation of an evidentiary DNA profile should not be influenced by information about a subject's DNA profile.** Each item of evidence must be interpreted independently of other items of evidence or reference samples. Yet forensic analysts are commonly aware of submitted reference profiles when interpreting DNA test results, creating the opportunity for confirmatory bias, despite the best intentions of the analyst…"

http://www.dfs.virginia.gov/about/minutes/saCommittee/20080108.pdf

DNA Mixture Interpretation:

Principles and Practice in Component Deconvolution and Statistical Analysis

Principles in Mixture Interpretation

Handouts available on STRBase at
http://www.cstl.nist.gov/biotech/strbase/training/AAFS2008_MixtureWorkshop.htm

AAFS 2008 Workshop #16
Washington, DC
February 19, 2008

John M. Butler

john.butler@nist.gov

Two Parts to Mixture Interpretation

- Determination of alleles present in the evidence and **deconvolution of mixture components** where possible
  - Many times through comparison to victim and suspect profiles
- **Providing some kind of statistical answer** regarding the weight of the evidence
  - There are multiple approaches and philosophies

Software tools can help with one or both of these…

Status of Software for Mixture Interpretation

- NIJ Expert System Testbed (**NEST**) Project
  - Evaluating software programs for DNA analysis of single-source (Phase I) and mixtures (Phase II)
  - **http://forensics.marshall.edu/NEST/NEST-Intro.html**
- US Army Crime Laboratory (**USACIL**)
  - Commonly deal with complex sexual assaults
  - Developed software for aiding mixture interpretation and statistical analysis

Steps in the Interpretation of Mixtures

(Clayton *et al.* 1998)

Step #1 Identify the Presence of a Mixture

Step #2 Designate Allele Peaks

Step #3 Identify the Number of Potential Contributors

Step #4 Estimate the Relative Ratio of the Individuals Contributing to the Mixture

Step #5 Consider All Possible Genotype Combinations

Step #6 Compare Reference Samples


Analysis and interpretation of mixed forensic stains using DNA STR profiling

Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70

Adapted from Peter Schneider slide (presented at EDNAP meeting in Krakow in April 2007)

Mixture Classification Scheme

Schneider *et al.* (2006) *Rechtsmedizin* 16:401-404

(German Stain Commission, 2006):

- **Type A**: no obvious major contributor, no evidence of stochastic effects
- **Type B**: clearly distinguishable major and minor contributors; consistent peak height ratios of ***approximately 4:1*** (major to minor component) for all heterozygous systems, no stochastic effects
- **Type C**: mixtures without major contributor(s), evidence for stochastic effects

Type A **Type B** **Type C**

Type of mixture and interpretation

- **Type A:** Mixed profile without stochastic effects, a biostatistical analysis has to be performed
- **Type B:** Profile of a major contributor can be unambiguously described and interpreted as a profile from an unmixed stain
- **Type C:** due to the complexity of the mixture, the occurrence of stochastic effects such as allele and locus drop-outs have to be expected:
  - a clear decision to include or exclude a suspect may be difficult to reach, thus a biostatistical interpretation is not appropriate.

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

Biostatistical approaches

- Calculation of the **probability of exclusion** for a randomly selected stain donor* [P(E)]
  (*RMNE - "random man not excluded")
- Calculation of the **likelihood ratio** [LR] based on defined hypotheses for the origin of the mixed stain

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

Which approach should be used?

- If the basis for clearly defined and mutually exclusive hypotheses is given, i.e.:
  - the number of contributors to the stain can be determined,
  - unambiguous DNA profiles across all loci are observed (type A mixtures, or type B, if the person considered as "unknown" contributor is part of the minor component of the mixture),

  then the calculation of a likelihood ratio is appropriate.

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

Which approach should be used?

- If major/minor contributors cannot be identified based on unambiguous DNA profiles, or if the the number of contributors cannot be determined, then the calculation of the probability of exclusion is appropriate.
- The calculation of P(E) is always possible for type A and type B mixtures.

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

Not acceptable …

- … is the inclusion of a genotype frequency of a non-excluded suspect into the report, if the given mixed stain does not allow a meaningful biostatistical interpretation.
  - this would lead to the wrongful impression that this genotype frequency has any evidentiary value regarding the role of the suspect as a contributor to the mixed stain in question.

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

Conclusions

- The likelihood ratio has a significant weight of evidence, as it relates directly to the role of the suspect in the context of the origin of the stain.
- The exclusion probability makes a general statement without relevance to the role of the suspect.
- However, this does not imply that P(E) is always more "conservative" in the sense that the weight of evidence is not as strong compared to the LR.

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

GEDNAP 32

Mixture interpretation exercise:

- 3 person mixture without major contributor
- Person A from group of reference samples was not excluded
- Allele frequencies for eight German database systems provided for exercise
- German-speaking GEDNAP participants invited to participate based on published recommendations

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

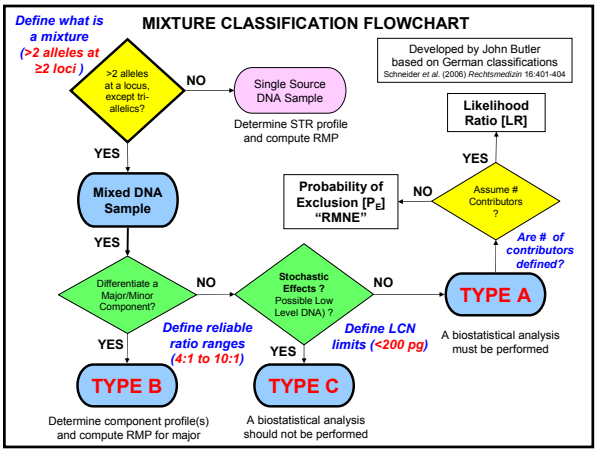
GEDNAP 32

Results:

- 22 labs submitted results (from approx. 80 German-speaking GEDNAP participants)
- Calculations submitted were all correct and consistent:
  - 15x LR approach:
    - Person A + 2 unknown vs. 3 unknown contributors
  - 11x RMNE calculation
- Will be offered again next time

Training and Specific Guidelines/Classification Schemes yielded consistent results among laboratories

Slide from Peter Schneider (presented at EDNAP meeting in Krakow in April 2007)

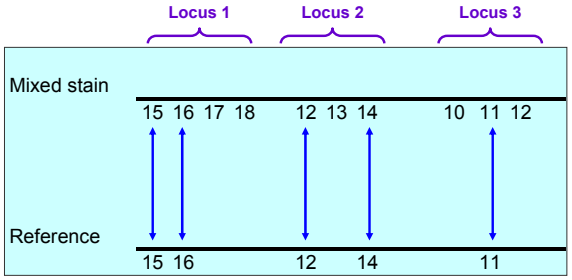


German Type A,B, and C mixture classifications

- **Type A**, where major/minor contributors cannot be deduced, require stats
  - LR
  - RMNE
- **Type B** enables major contributor to be deduced
  - RMP (which is 1/LR)
- **Type C** no stats should be attempted because of the possibility of failure to account for allele dropout due to stochastic effects with low level DNA samples

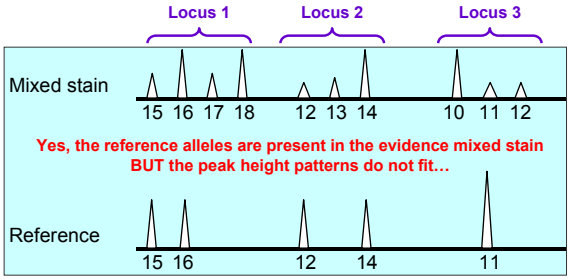
Mixture Example

Comparing Alleles Only



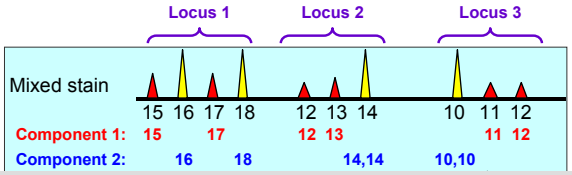
Mixture Example

Showing Importance of Using Peak Height Information



Mixture Example

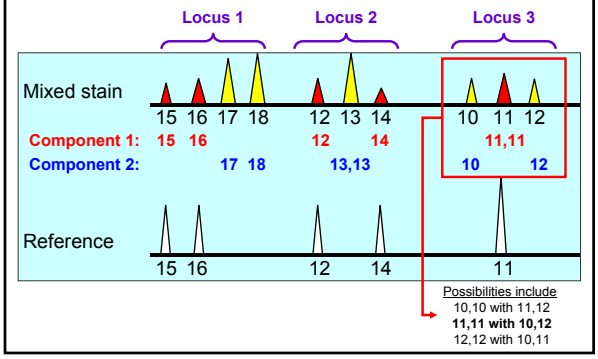
Solving Components Prior to Comparison to Suspect Reference



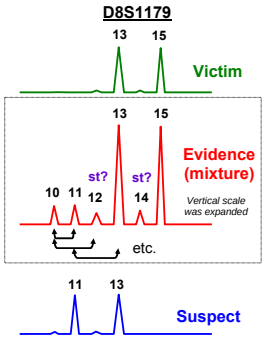
Reference (suspect) does not match either component of the mixed stain and therefore could not have contributed to the evidence sample

Mixture Example

Different Evidence Sample…



Another Mixture Example



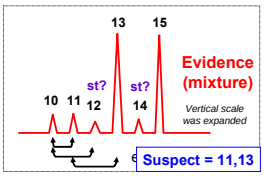
Conclusions from the evidence:

1. **Major contributor = 13,15 (victim) –** *to be expected with an intimate sample like a fingernail or vaginal swab*
2. Alleles 12 and 14 are **likely stutter products** of the major contributor's 13 and 15 alleles but could also be masking minor contributor alleles
3. A number of minor contributor combinations are possible (e.g., 10,11 or 10,12 or 10,13 or 11,13, etc.)
4. Could have more than two contributors present in this mixture

"Suspect cannot be excluded" BUT statement needs to be qualified by statistics because a large percentage of the population might also not be able to be excluded…

Probability of Exclusion Calculation
for a Single STR Locus

The case may grow stronger against a suspect with information from additional STR loci…



From VA DFS STR Allele Frequencies
http://www.dfs.virginia.gov/manuals/manuals.cfm?id=5

D8S1179 alleles	AA (n=384)	C (n=346)	H (n=366)
10	0.0287	0.1069	0.0820
11	0.0495	0.0925	0.0465
12	0.1094	0.1416	0.1093
13	0.2422	0.3093	0.3224
14	0.2969	0.1965	0.2623
15	0.1849	0.0896	0.1202
SUM	0.9115	0.9364	0.9426
Sq SUM = PI	0.8308	0.8769	0.8886
PE = 1-PI	0.1692	0.1231	0.1114
PE (%)	**16.9%**	**12.3%**	**11.1%**
	African Am.	Caucasians	Hispanics

"Suspect cannot be excluded" BUT we would expect to see, for example, only 11.1% of Hispanics excluded (or 88.9% cannot be excluded) based on results at this one locus

The fact that in this case a suspect is included is not very informative because ~9 out of 10 people examined from any population could potentially be included in the evidence mixture…

The Statistic (Determining the Weight of the Evidence) Should Be **Calculated from the Evidence**

Evidence (partial profile):

	Type	Statistic
Locus 1	16,17	1 in 9
Locus 2	17,18	1 in 9
Locus 3	21,22	1 in 12
Locus 4	12,14	1 in 16
Locus 5	28,30	1 in 11

		Product = 1 in 171,000

Match Observed at All Loci that May Be Compared

Reference (full profile):

	Type	Statistic
Locus 1	16,17	1 in 9
Locus 2	17,18	1 in 9
Locus 3	21,22	1 in 12
Locus 4	12,14	1 in 16
Locus 5	28,30	1 in 11
Locus 6	14,16	1 in 26
Locus 7	12,13	1 in 9
Locus 8	11,14	1 in 31
Locus 9	9,9	1 in 32
Locus 10	9,11	1 in 14
Locus 11	6,6	1 in 19
Locus 12	8,8	1 in 3
Locus 13	10,10	1 in 21

		Product = 1 in 665 trillion

The reference sample is still a "match" – just not as much information is available from the evidence for comparison

Statistical Approaches with Mixtures

See Ladd *et al.* (2001) *Croat Med J.* 42:244-246

- **Inferring Genotypes of Contributors** - Separate major and minor components into individual profiles and compute the random match probability estimate as if a component was from a single source
- **Calculation of Exclusion Probabilities** - CPE/CPI (RMNE) – The probability that a random person (unrelated individual) would be excluded as a contributor to the observed DNA mixture
- **Calculation of Likelihood Ratio Estimates** – Comparing the probability of observing the mixture data under two (or more) alternative hypotheses; in its simplest form LR = 1/RMP

RMNE = Random Man Not Excluded (same as CPE)
CPE = Combined Probability of Exclusion (CPE = 1 – CPI)
CPI = Combined Probability of Inclusion (CPI = 1 – CPE)

Advantages and Disadvantages

RMNE (CPE/CPI)

- **Advantages**
  - Does not require an assumption of the number of contributors to a mixture
  - Easier to explain in court
- **Disadvantages**
  - Weaker use of the available information (robs the evidence of its true probative power because this approach does not consider the suspect's genotype)
  - Likelihood ratio approaches are developed within a consistent logical framework

Likelihood Ratios (LR)

- **Advantages**
  - Enables full use of the data including different suspects
- **Disadvantages**
  - More difficult to calculate

John Buckleton, *Forensic DNA Evidence Interpretation*, p. 223

Assumptions for CPE/CPI Approach

- There is no allele dropout (i.e., all alleles are above stochastic threshold) – low-level mixtures can not reliably be treated with CPE
- All contributors are from the same racial group (i.e., you use the same allele frequencies for the calculations)
- All contributors are unrelated
- Peak height differences between various components are irrelevant (i.e., component deconvolution not needed) – this may not convey all information from the available sample data…

Likelihood Ratio (LR)

- Provides ability to express and evaluate both the prosecution hypothesis, H_p (the suspect is the perpetrator) and the defense hypothesis, H_d (an unknown individual with a matching profile is the perpetrator)

$$LR = \frac{H_p}{H_d}$$

- **The numerator, H_p, is usually 1** – since in theory the prosecution would only prosecute the suspect if they are 100% certain he/she is the perpetrator
- The denominator, $\mathbf{H_d}$, is typically the profile frequency in a particular population (based on individual allele frequencies and assuming HWE) – i.e., **the random match probability**

LR is not a probability but a ratio of probabilities

DAB Recommendations on Statistics

February 23, 2000
Forensic Sci. Comm. 2(3); available on-line at
http://www.fbi.gov/hq/lab/fsc/backissu/july2000/dnastat.htm

"The DAB finds either one or both PE or LR calculations acceptable and strongly recommends that one or both calculations be carried out whenever feasible and a mixture is indicated"

- Probability of exclusion (PE)
  - Devlin, B. (1993) Forensic inference from genetic markers. *Statistical Methods in Medical Research*, 2, 241–262.
- Likelihood ratios (LR)
  - Evett, I. W. and Weir, B. S. (1998) *Interpreting DNA Evidence*. Sinauer, Sunderland, Massachusetts.

ISFG DNA Commission on Mixture Interpretation

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

Available for download from the ISFG Website:
http://www.isfg.org/Publication;Gill2006


Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Forensic Science International

DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

Summary of ISFG Recommendations on Mixture Interpretation

1. The likelihood ratio (LR) is the preferred statistical method for mixtures over RMNE
2. Scientists should be trained in and use LRs
3. Methods to calculate LRs of mixtures are cited
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
5. Prosecution determines H_p and defense determines H_d and multiple propositions may be evaluated
6. When minor alleles are the same size as stutters of major alleles, then they are indistinguishable
7. Allele dropout to explain evidence can only be used with low signal data
8. No statistical interpretation should be performed on alleles below threshold
9. Stochastic effects limit usefulness of heterozygote balance and mixture proportion estimates with low level DNA

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

Thoughts by Peter Gill on Recommendation #5

(ENFSI meeting, Krakow, Poland, April 19, 2007)

- Prosecution and defense each want to maximize their respective probabilities
- Recommendation 5 places ownership for each hypothesis.
- In order to perform the LR calculation(s), the forensic scientist decides on both the prosecution and defense hypotheses.
- Since the forensic scientists usually cannot discover the defense hypothesis before the trial (as they are typically working with the prosecution if the DNA matches…), assumptions must be clearly stated with the important caveat that you cannot perform calculations on the stand! (For example, you need three weeks warning to make and check calculations.)
- By anchoring the respective hypotheses to each side, the defense can change their hypothesis but the prosecution does not need to change theirs…
- It is worth noting that the likelihood ratio always goes up if the defense lowers their hypothesis (H_d gets lower with more possible combinations)

ISFG (2006) Recommendations

- Recommendation 6: If the crime profile is a major/minor mixture, **where minor alleles are the same size (height or area) as stutters of major alleles, then stutters and minor alleles are indistinguishable.** Under these circumstances alleles in stutter positions that do not support H_p should be included in the assessment.
- In general, stutter percentage is <15%

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

Consideration of Peak in Stutter Position

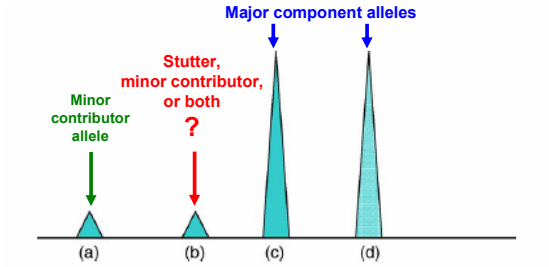


Fig. 4. *c* and *d* are unambiguous alleles, *b* is a minor allele in a stutter position and *a* is an unambiguous minor allele.

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

UK Response

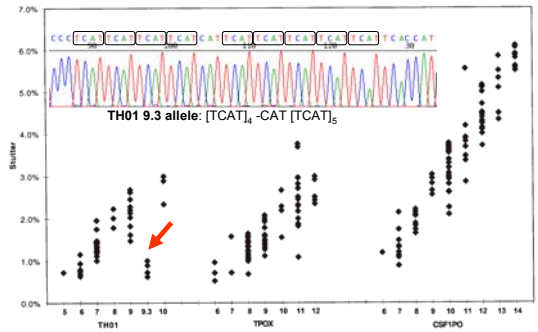
Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

Recommendation 6:

- Stutters are locus-dependent…
- It is recommended that laboratories make their own maximum experimentally observed stutter sizes per locus determinations since the effects may be technique dependent.
- It is recommended that [maximum stutter percentages be] evaluated per locus.

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.

UK Response

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

- ***Characterization of +4 base stutters***

We agreed to review +4 bp stutters, however, we note that their presence often relates to over-amplified samples. Preliminary experimental work suggests that they are low level and **generally less then 4% the size of the progenitor allele** (Rosalind Brown, personal communication). Note that 4 bp and +4 bp stutter cannot be distinguished from genetic somatic mutation without experimental work—furthermore, somatic mutations may give rise to peaks that are larger than those caused by stutter artifacts.

ISFG (2006) Recommendations

- Recommendation 7: If drop-out of an allele is required to explain the evidence under H_p: (S = ab; E = a), then the allele should be small enough (height/area) to justify this. Conversely, if a full crime stain profile is obtained where alleles are well above the background level, and the probability of drop-out approaches Pr(D) ≈ 0, then H_p is not supported.

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

UK Response

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

Recommendation 7:

- We recommend slight rewording…[with mention of companion allele]
- If a full crime-stain profile is obtained where alleles are well above the background level, and the probability of dropout Pr(D) approaches zero, then H_p is not supported (Figure 6).

Hypothetical Examples

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

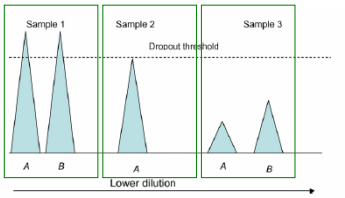


Fig. 4. Results from serial dilutions of the same sample genotype *AB*. The first result (sample 1) shows a locus where both alleles are represented in the profile. One or both of these alleles are above the dropout threshold and consequently are always present in the epg. The second result shows a result where dropout has occurred – the survivor allele is just below the dropout threshold hence this is a rare event, but not impossible. If *A* was just above the dropout threshold we would determine it to be a homozygote *AA* genotype. In the third sample, both alleles are well below the dropout threshold – it is an unambiguous, albeit unbalanced heterozygote. If only one allele was present, then we would have to consider the possibility of dropout of the partner. The same rationale can be applied to any analytical regime, e.g. 28 and 34 PCR cycles.

If Below Dropout Threshold…

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

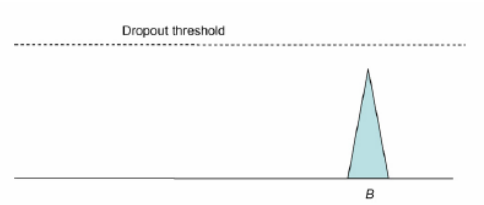


Fig. 5. In this example allele *B* is below the dropout threshold, hence we cannot be confident that it is from a homozygote *BB* individual. It could also be from an individual who is heterozygote, where the missing allele is any other allele. The probability *B*|unknown, H_d is 2Pr(*BF*), where the '*F*' designation is assigned a probability of 1 to take account of the possibility that any allele could have dropped out.

If Above Dropout Threshold…

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

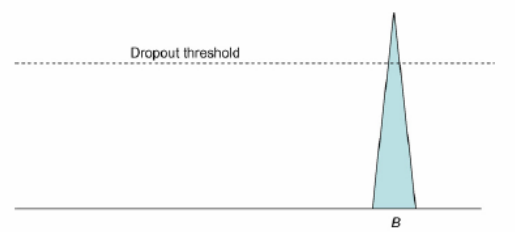


Fig. 6. In this example allele *B* is above the dropout threshold, hence we can be confident that it is from a homozygote *BB* individual. The probability of B|unknown, H_d is $\Pr(B)^2$.

Setting Thresholds

- **Detection (analytical) threshold**
  - Dependent on instrument sensitivity
  ~50 RFU
  - Impacted by instrument baseline noise
- **Dropout (stochastic) threshold**
  - Dependent on biological sensitivity
  ~150-200 RFU
  - Impacted by assay and injection parameters

Determining the Dropout (Stochastic) Threshold

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

- The dropout threshold can be determined experimentally for a given analytical technique from a series of pre-PCR dilutions of extracts of known genotype technique (it will probably vary between analytical methods). These samples can be used to determine the point where allelic dropout of a heterozygote is observed relative to the size of the survivor companion allele. The threshold is the maximum size of the companion allele observed. This is also the point where Pr(D) approaches zero (Fig. 4).

Dropout threshold will change depending on instrument and assay conditions (e.g., longer CE injection will raise dropout threshold)

ISFG (2006) Recommendations

- Recommendation 8: If the alleles of certain loci in the DNA profile are at a level that is dominated by background noise, then a biostatistical interpretation for these alleles should not be attempted.

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

UK Response

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

Recommendation 8:

- If there is a band below the experimental threshold where background noise might be prevalent, and it is distinct and clear from the background, then it should be recorded and available on the case file.

ISFG (2006) Recommendations

- Recommendation 9: In relation to low copy number, stochastic effects limit the usefulness of heterozygous balance and mixture proportion estimates. In addition, allelic drop-out and allelic drop-in (contamination) should be taken into consideration of any assessment.

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

UK Response

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

Recommendation 9:

- Case pre-assessment is necessary in order to determine the best scientific method to process a sample. To facilitate this, it is recommended that wherever possible, this should include quantification. Quantification is used to determine the optimum method to process—if low-level DNA, a sample would benefit from procedures to enhance sensitivity of detection. There may be reasons where quantification is not practicable, especially if low levels of DNA are expected, since the result itself may be compromised if a portion of the sample is sacrificed. At low DNA levels, the accuracy of the quantification test itself may be inefficient.

UK Response

Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

Recommendation 9 (cont):

- It is possible that a given DNA profile may simultaneously comprise both 'conventional' and 'low-level' loci: for example, if degradation has occurred then low molecular weight loci may be above the dropout threshold, whereas high molecular weight loci may be below the dropout threshold.
- Similarly, if the sample is a mixture, then at a given locus there may be some alleles that are above the dropout threshold (from a major contributor) and others that are below the dropout threshold (from a minor contributor), i.e. different interpretation rationale may be simultaneously applied to different contributors within a locus.

Thank you for your attention…

Questions
or Comments?

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

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

Summary of ISFG Recommendations on Mixture Interpretation

1. The likelihood ratio (LR) is the preferred statistical method for mixtures over RMNE
2. Scientists should be trained in and use LRs
3. Methods to calculate LRs of mixtures are cited
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
5. Prosecution determines H_p and defense determines H_d and multiple propositions may be evaluated
6. When minor alleles are the same size as stutters of major alleles, then they are indistinguishable
7. Allele dropout to explain evidence can only be used with low signal data
8. No statistical interpretation should be performed on alleles below threshold
9. Stochastic effects limit usefulness of heterozygote balance and mixture proportion estimates with low level DNA

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

MIXTURE CLASSIFICATION FLOWCHART

Define what is a mixture (>2 alleles at ≥2 loci)

Developed by John Butler
based on German classifications
Schneider *et al.* (2006) *Rechtsmedizin* 16:401-404

>2 alleles at a locus, except tri-allelics?

NO

Single Source DNA Sample

Determine STR profile and compute RMP

YES

Mixed DNA Sample

YES

Differentiate a Major/Minor Component?

NO

Stochastic Effects ? Possible Low Level DNA) ?

Define reliable ratio ranges (4:1 to 10:1)

Define LCN limits (<200 pg)

YES

TYPE B

Determine component profile(s) and compute RMP for major

YES

TYPE C

A biostatistical analysis should not be performed

NO

TYPE A

A biostatistical analysis must be performed

Are # of contributors defined?

Assume # Contributors ?

YES

Likelihood Ratio [LR]

NO

Probability of Exclusion [P_E] "RMNE"

Mixture Deconvolution

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

Outline

- Points for Consideration
  - DNA quantity and quality
- Deconvolution steps by Clayton *et al.* (1998)
- Worked Example – using NEST data
- Software programs introduced

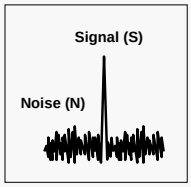
Final version available at
http://www.cstl.nist.gov/biotech/strbase/training/AAFS2008_MixtureWorkshop.htm

Points for Consideration

- Peak height vs peak area
- Thresholds – analytical vs stochastic levels
- Other lab-specific values:
  - Heterozygote peak height balance
  - Locus-specific stutter percentage
- DNA quantity and quality
  - problems with low-level or degraded DNA

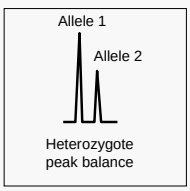
What is a true peak (allele)?

Peak detection threshold



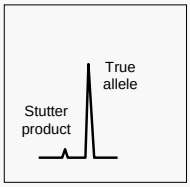
Signal > 3x sd of noise

Peak height ratio (PHR)



PHR consistent with single source Typically above 60%

Stutter percentage



Stutter location below 15%

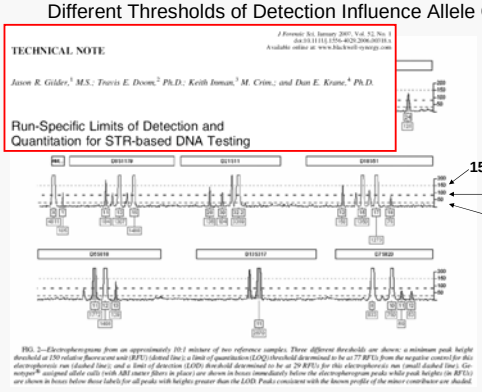
Validation Studies

- Information from validation studies should be used to set laboratory-specific
  - Stutter %
  - Peak Height Ratios
  - Minimum Peak Heights (detection thresholds)
  - Relative balance across loci
- These values are all dependent on amount of input DNA
  - If low-level DNA is amplified, stutter % may be higher and peak height ratios may be lower

Thresholds

- Validation studies should be performed in each laboratory
- Some labs have set two thresholds:
  - Analytical thresholds – what is a peak? (50 RFU)
  - Stochastic thresholds – what is reliable PCR data? (150 RFU)

Different Thresholds of Detection Influence Allele Calls

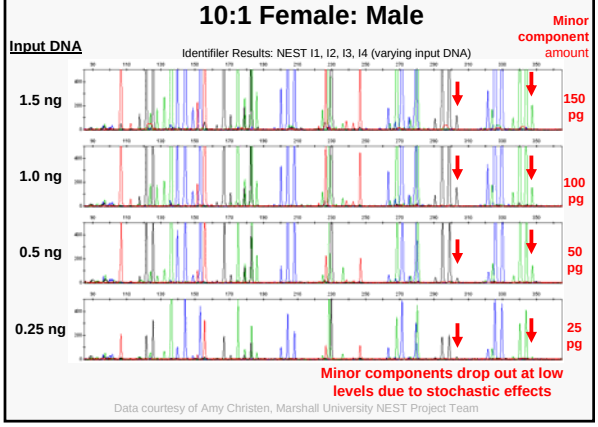


Gilder, J.R., Doom, T.E., Inman, K., Krane, D.E. (2007) Run-specific limits of detection and quantitation for STR-based DNA testing. *J. Forensic Sci.* 52(1): 97-101.

The Scientific Reasoning behind the Concept of an Analytical Threshold (limit of detection)

- This is fundamentally an issue of reliability
- **For a peak intensity three times the standard deviation of the noise there is a limited chance that such a signal is the result of a random fluctuation**
- This is because 99.7 percent of all noise signals fall below this value (from the definition of a Gaussian curve)
- **Below this point the very real possibility exists that what you think is a peak is simply a statistical fluctuation in the baseline noise**.

10:1 Female: Male



Steps in the Interpretation of Mixtures

(Clayton *et al.* 1998)

Step #1 Identify the Presence of a Mixture

Step #2 Designate Allele Peaks

Step #3 Identify the Number of Potential Contributors

Step #4 Estimate the Relative Ratio of the Individuals Contributing to the Mixture

Step #5 Consider All Possible Genotype Combinations

Step #6 Compare Reference Samples

Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70

Will review each step with a worked example

Step #1: Is a Mixture Present in an Evidentiary Sample?

- Examine the number of peaks present in a locus
  - More than 2 peaks at a locus (except for tri-allelic patterns at perhaps one of the loci examined)
- Examine relative peak heights
  - Heterozygote peak imbalance <60%
  - Peak at stutter position >15%
- Consider all loci tested

Is a DNA Profile Consistent with Being a Mixture?

From J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*, pp. 156-157

If the answer to any one of the following three questions is yes, then the DNA profile may very well have resulted from a mixed sample:

- Do any of the loci show more than two peaks in the expected allele size range?
- Is there a severe peak height imbalance between heterozygous alleles at a locus?
- Does the stutter product appear abnormally high (e.g., >15-20%)?

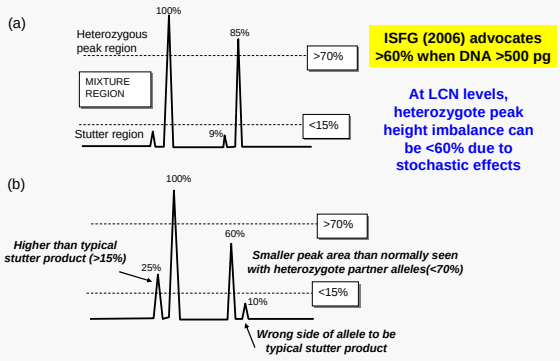


Figure 7.3, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Step #2: Designate Allele Peaks

- Use regular data interpretation rules to decipher between true alleles and artifacts
- Use stutter filters to eliminate stutter products from consideration (although stutter may hide some of minor component alleles at some loci)
- Consider heterozygote peak heights that are highly imbalanced (<60%) as possibly coming from two different contributors

Step #3: Identifying the Potential Number of Contributors

- **Important for some statistical calculations**
- Typically if 2, 3, or 4 alleles then 2 contributors
- If 5 or 6 alleles per locus then 3 contributors
- If >6 alleles in a single locus, then >4 contributors

- JFS Nov 2005 paper by Forensic Bioinformatics on number of possible contributors
  - Relies on maximum allele count alone
  - Does not take into account peak height information

Forensic Bioinformatics Article

http://www.bioforensics.com/articles/empirical_mixtures.pdf

J Forensic Sci, Nov. 2005, Vol. 50, No. 6
Paper ID JFS2004475
Available online at: www.astm.org

David R. Paoletti,[1] M.S.; Travis E. Doom,[1,2] Ph.D.; Carissa M. Krane,[3] Ph.D.; Michael L. Raymer,[1,2] Ph.D.; and Dan E. Krane,[4] Ph.D.

Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures

TABLE 2—*Count and percent of three-person mixtures in which a particular number of unique alleles was the maximum observed across all loci, both for the original and randomized individuals*[a].

Unique Alleles	Count	Percent (%)
2	0	0.00%
3	78	0.00%
4	4,967,034	3.39%
5	93,037,010	63.49%
6	48,532,037	33.12%

Using 959 complete 13-locus STR profiles from FBI dataset

146,536,159 possible combinations **with 3-person mixtures**

3.39 % (4,967,034 combinations) would only show a maximum of four alleles (i.e., appear based on maximum allele count alone to be a 2-person mixture)

Recent Article by Buckleton *et al.*

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 1 (2007) 20–28
FSI GENETICS
www.elsevier.com/locate/fsig

Towards understanding the effect of uncertainty in the number of contributors to DNA stains

John S. Buckleton[a], James M. Curran[b,*], Peter Gill[c]
[a]The Institute of Environmental Science and Research Ltd., Private Bag 92021, Auckland, New Zealand
[b]Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand
[c]The Forensic Science Service, Trident Court, Solihull Parkway, Birmingham Business Park, Solihull B37 7YN, UK

Received 31 May 2006; received in revised form 12 September 2006; accepted 13 September 2006

Abstract

DNA evidence recovered from a scene or collected in relation to a case is generally declared as a mixture when more than two alleles are observed at several loci. However, in principle, all DNA profiles may be considered to be potentially mixtures, even those that show not more than two alleles at any locus. When using a likelihood ratio approach to the interpretation of mixed DNA profiles it is necessary to postulate the number of potential contributors. However, this number is never known with certainty. The possibility of a, say three-person mixture, presenting four or fewer peaks at each locus of the CODIS set was explored by Paoletti et al. [D.R. Paoletti, T.E. Doom, C.M. Krane, M.L. Raymer, D.E. Krane, Empirical analysis of the STR profiles resulting from conceptual mixtures, J. Forensic Sci. 50 (2005) 1361–1366]. In this work we extend this analysis to consider the profiler plus and SGM plus multiplices. We begin the assessment of the risk associated with current practice in the calculation of LR's. We open the discussion of possible ways to surmount this ambiguity.

Two-Person Mixtures for Simulated Profiles:

Probability by Locus of A Particular Number of Alleles Being Observed

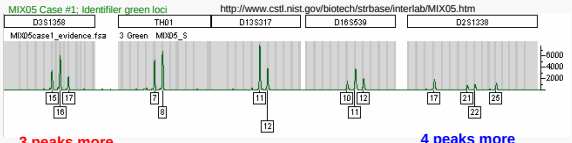
Table 1
The probability of observing a given number of alleles in a two-person mixtures for simulated profiles at the SGM+™ loci

Loci	No. of alleles			
	1	2	3	4
D3	0.011	0.240	0.559	0.190
vWA	0.008	0.194	0.548	0.250
D16	0.016	0.287	0.533	0.164
D2	0.003	0.094	0.462	0.441
D8	0.011	0.194	0.521	0.274
D21	0.007	0.147	0.505	0.341
D18	0.003	0.095	0.472	0.430
D19	0.020	0.261	0.516	0.203
THO	0.016	0.271	0.547	0.166
FGA	0.003	0.116	0.500	0.381

Buckleton et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28

Levels of Locus Heterozygosity Impact Number of Alleles Observed in Mixtures

Loci	No. of alleles			
	1	2	3	4
D3	0.011	0.240	0.559	0.190
vWA	0.008	0.194	0.548	0.250
D16	0.016	0.287	0.533	0.164
D2	0.003	0.094	0.462	0.441



3 peaks more common for D3

4 peaks more common for D2

Three-Person Mixtures for Simulated Profiles:

Probability by Locus of A Particular Number of Alleles Being Observed

Table 2
The probability of observing a given number of alleles in a three-person mixtures for simulated profiles at the SGM+™ loci

Loci	No. of alleles showing					
	1	2	3	4	5	6
D3	0.000	0.053	0.366	0.463	0.115	0.002
vWA	0.000	0.037	0.285	0.468	0.194	0.016
D16	0.001	0.086	0.397	0.411	0.100	0.005
D2	0.000	0.008	0.104	0.385	0.393	0.110
D8	0.001	0.041	0.258	0.436	0.236	0.029
D21	0.000	0.023	0.192	0.428	0.302	0.055
D18	0.000	0.007	0.109	0.392	0.396	0.096
D19	0.003	0.078	0.352	0.401	0.152	0.014
THO	0.001	0.074	0.395	0.439	0.088	0.002
FGA	0.000	0.012	0.144	0.424	0.346	0.074

Buckleton et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28

Number of Alleles Observed with Simulated Four-Person Mixtures

- The simulation of four person mixtures suggests that 0.014% of four person mixtures would show four or fewer alleles and that 66% would show six or fewer alleles for the SGM Plus loci.
- The results for the Profiler Plus loci were 0.6% and 75%.
- The equivalent values for the CODIS set from Paoletti et al. were 0.02% showing four or fewer and 76.35% showing six or fewer.

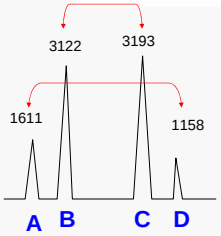
Buckleton et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28

Step #4: Estimation of Relative Ratios for Major and Minor Components to a Mixture

- Mixture studies with known samples have shown that the mixture ratio between loci is fairly well preserved during PCR amplification
- Thus it is generally thought that the peak heights (areas) of alleles present in an electropherogram can be related back to the initial component concentrations
- Start with loci possessing 4 alleles…

Estimating Mixture Proportion (M_x) or Mixture Ratio (M,)

$$M_x = \frac{\phi_a + \phi_b}{\phi_a + \phi_b + \phi_c + \phi_d}$$



$$\frac{A + D}{A + D + B + C}$$

$$\frac{1611 + 1158}{1611 + 1158 + 3122 + 3193}$$

= 2769/9084 = 0.305

Step #5: Consider All Possible Genotype Combinations

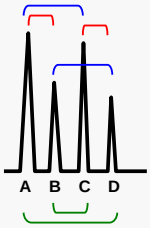
Table 3
Pairwise combinations of two, three and four alleles

Four alleles (a,b,c,d)		Three alleles (a,b,c)		Two alleles (a,b)	
a,b	c,d	a,a	b,c	a,a	a,b
a,c	b,d	b,b	a,c	a,b	a,b
a,d	b,c	c,c	a,b	a,a	b,b
c,d	a,b	a,b	a,c	a,b	b,b
b,d	a,c	b,c	a,c	a,b	a,a
b,c	a,d	a,b	b,c	b,b	a,a
		b,c	a,a	b,b	a,b
		a,c	b,b		
		a,b	c,c		
		a,c	a,b		
		a,c	b,c		
		b,c	a,b		

Key: bold entries represent reciprocal combinations.

Clayton *et al. Forensic Sci. Int.* 1998; 91:55-70

Considering Genotype Combinations



AC
BD
AB
CD
BC
AD

Depends on PHR

Peak Height Ratios (PHR)
Minimum Peak Height (mPH)
Proportion (p) or mixture proportion (M_x)

Step #6: Compare Reference Samples

- If there is a suspect, a laboratory must ultimately decide to include or exclude him…
- **If no suspect is available for comparison, does your laboratory still work the case?** (Isn't this a primary purpose of the national DNA database?)
- Victim samples can be helpful to eliminate their allele contributions to intimate evidentiary samples and thus help deduce the perpetrator

Worked Example

NIJ Expert Systems Testbed (NEST) Project

http://www.promega.com/profiles/1002/ProfilesInDNA_1002_13.pdf

Profiles in DNA (September 2007) 10(2): 13-15 PROFILES IN DNA

EXPERT SYSTEMS

Validating Expert Systems: Examples with the FSS-i³™ Expert Systems Software

By Rhonda K. Roby[1] and Amy D. Christen[2]
[1]Technical Consultant, National Institute of Justice
[2]Research Analyst, Marshall University

NIJ National Institute of Justice

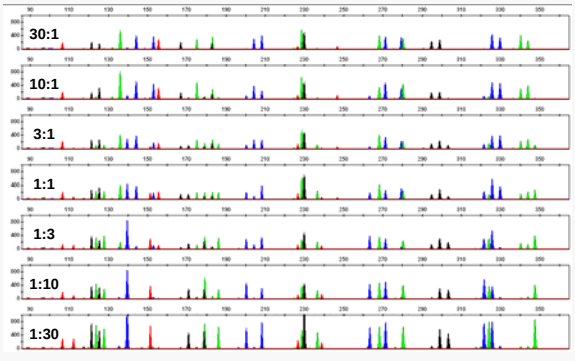
MARSHALL UNIVERSITY FORENSIC SCIENCE CENTER

NEST Project Mixture Sample Set

- NIJ Expert Systems Testbed (NEST) Project
  - Marshall University with Rhonda Roby (NIJ consultant)
- Phase II Mixture Sample Analysis
  - **Amy Christen** (Marshall University) produced a dataset while interning at Forensic Science Service in Summer 2006
  - Data to be used for evaluating "expert systems"
- Mixtures tested (280 total samples)
  - **2 different female/male sample combinations**: A:X and B:Y
  - **4 input DNA amounts**: 1.5 ng, 1.0 ng, 0.5 ng, 0.25 ng
  - **5 kits**: Identifiler, ProfilerPlus, COfiler, PowerPlex 16, SGM Plus
  - **7 mixture ratios**: 30:1, 10:1, 3:1, 1:1, 1:3, 1:10, 1:30

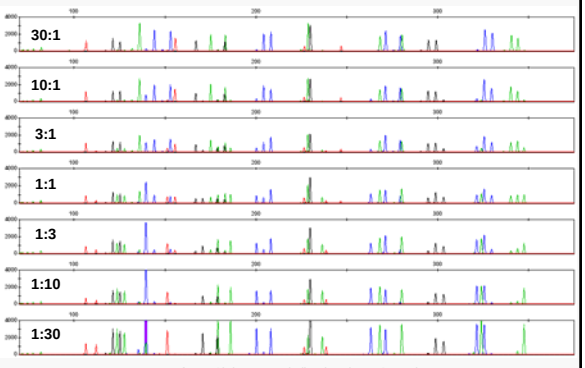
I will focus on a subset of this data… e.g., B:Y, 1.0 ng, Identifiler, 3:1

Identifiler Results: NEST H4 – N4 (**0.25 ng input DNA**)



Data courtesy of Amy Christen, Marshall University NEST Project Team

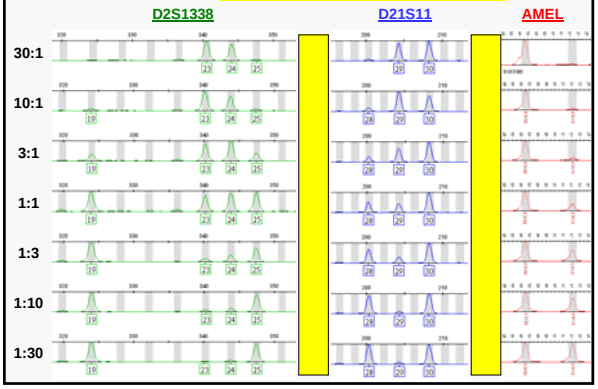
Identifiler Results: NEST H2 – N2 (**1.0 ng input DNA**)



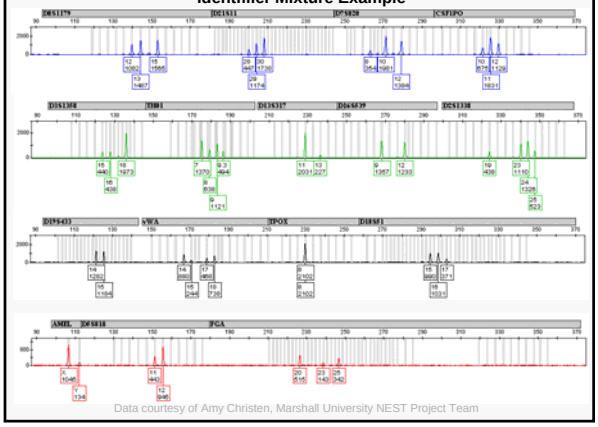
Data courtesy of Amy Christen, Marshall University NEST Project Team

Identifiler Results: NEST H2 – N2 (1.0 ng input DNA)

Calculate ratios based on peak heights

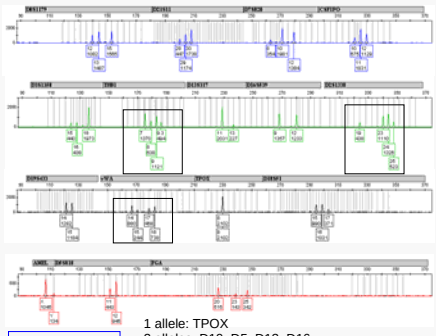


Identifiler Mixture Example



Data courtesy of Amy Christen, Marshall University NEST Project Team

Profile Overview



Evaluation Notes:

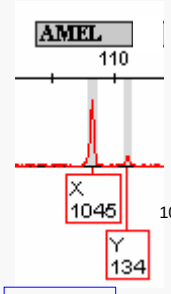
1. Loci seen with 1,2,3,&4 alleles (**a mixture** with at least 2 contributors)
2. Imbalance at amelogenin (female & male mixture with female as major)
3. Decent overall signal with D8 in ~1500 RFU (out of stochastic range)
4. Large MW loci have decent signal with D18 in ~1000 RFU range (degradation unlikely)
5. Ratio of **major to minor around 3:1** (from amelogenin X/Y ratios)

1045/134 = 7.8
~3 female (X,X):
1 male (X,Y)

1 allele: TPOX
2 alleles: D19, D5, D13, D16
3 alleles: D8, D21, D7, CSF, D3, D18, FGA
4 alleles: TH01, D2, VWA

Amelogenin Ratio

In many cases, amelogenin provides a helpful guide to assessing the mixture ratio



Female/Male ratio = X:X / X:Y

X/3 = 1045/3 = 348

348/134 = 2.6 (closest to 3 parts female to 1 part male)

1045/134 = 7.80

Chart of Expected Ratios

F:M	Chr ratio
1:1	3X:1Y
2:1	5X:1Y
3:1	7X:1Y
4:1	9X:1Y

1045/134 = 7.8
~3 female (X,X):
1 male (X,Y)

Potential problems with X or Y amplicon deletions

Anomalous Amelogenin Alleles

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

- **Males possessing only a single X amelogenin amplicon (Y null)** - *a male DNA sample will falsely look like a female DNA sample*:
  - Santos et al. (1998) reported a rare deletion of the amelogenin gene on the Y-chromosome
  - Y-STR typing can be performed to verify that other portions of the Y-chromosome are present
- **Males possessing only a single Y amelogenin amplicon (X null)**:
  - Shewale et al. (2000) observed loss of the X chromosome amplicon in three our of almost 7,000 males examined
  - while this phenomenon should not result in a gender misclassification (as the Y null situation might), its occurrence can impact the expected X and Y amplicon ratios in a mixture (see NIST MIX05 interlab study, case #3)

Running reference samples from suspect and/or victim may help discover potential amelogenin anomalies

Locus-by-Locus Breakdown…

- Start with 4 allele loci…
  - Assume two person mixture
  - With non-overlapping heterozygotes
  - Pair peaks with similar peak heights

Possible but not as likely depending on ratios

Possible Genotype Combinations

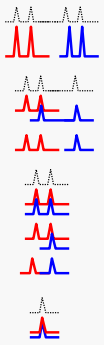
See Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, pp. 156-157

Four Peaks **(4 allele loci)**
- heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

Two Peaks **(2 allele loci)**
- heterozygote + heterozygote, two overlapping alleles (genotypes are identical)
- heterozygote + homozygote, one overlapping allele
- homozygote + homozygote, no overlapping alleles (genotypes are unique)

Single Peak **(1 allele loci)**
- homozygote + homozygote, overlapping allele (genotypes are identical)



MUST ALSO CONSIDER STUTTER POSITION

Population Database Used for STR Allele Frequencies

- U.S. population data contained in J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*, Appendix II (pp. 577-583)
- Published in Butler *et al.* (2003) *J. Forensic Sci.* 48(4): 908-911
- Available at http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm
- Will focus on Caucasians for simplicity

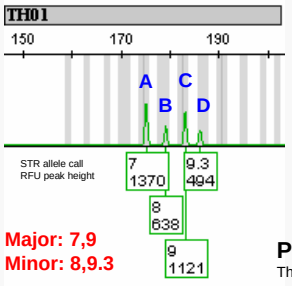
TH01

Allele	Caucasian N = 302	African-American N = 258	Hispanic N = 140
5	0.00166*	0.00388*	
6	0.23179	0.12403	0.21429
7	0.19040	**0.42054**	0.27857
8	0.08444	0.19380	0.09643
9	0.11424	0.15116	0.15000
9.3	**0.36755**	0.10465	0.24643
10	0.00828	0.00194*	0.01429*
11	0.00166*		

Remember that different population databases will have different allele frequencies because they are based on different samples

Stats

4 Allele Locus: TH01



Allele	Frequency
7	0.190
8	0.084
9	0.114
9.3	0.368

$PI = (P_A + P_B + P_C + P_D)^2$
$= (0.190 + 0.084 + 0.114 + 0.368)^2$
$= (0.756)^2$
$= 0.572$

Major: 7,9
Minor: 8,9.3

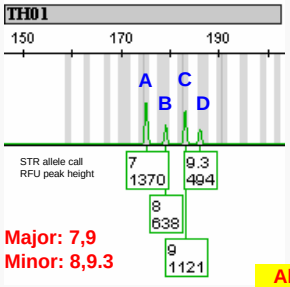
PE = 1 – PI = 1 – 0.572 = 0.428

Thus ~43% of Caucasian population can be excluded from contributing to this mixture (primarily because allele 6 is missing)

Four Peaks **(4 allele loci)**
- heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

PHRs

4 Allele Locus: TH01



Major: 7,9
Minor: 8,9.3

Consider all possible combinations:

B/A = 638/1370 = 0.466

B/C = 638/1121 = 0.569

C/A = 1121/1370 = 0.818 **major**

D/B = 494/648 = 0.774 **minor**

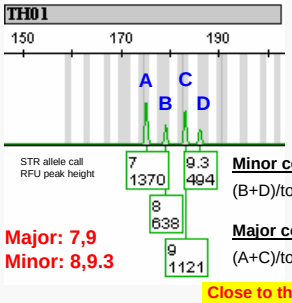
D/C = 494/1121 = 0.441

All other combinations <0.60 (60% heterozygote Peak Height Ratio)

Four Peaks **(4 allele loci)**
- heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

Mix Ratio

4 Allele Locus: TH01



Major: 7,9
Minor: 8,9.3

Total of all peak heights
= 1370 + 638 + 1121 + 494
= 3623 RFUs

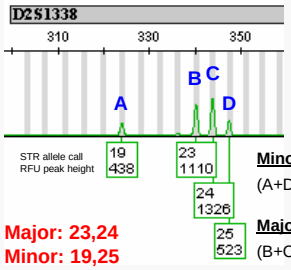
Minor component:
(B+D)/total = (638+494)/3623 = **0.312**

Major component:
(A+C)/total = (1370+1121)/3623 = **0.688**

Close to the ~3:1 predicted by amelogenin X/Y allele ratio – thus major component = female

Four Peaks **(4 allele loci)**
- heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

Mix Ratio
4 Allele Locus: D2S1338



Total of all peak heights
= 438 + 1110 + 1326 + 523
= 3397 RFUs

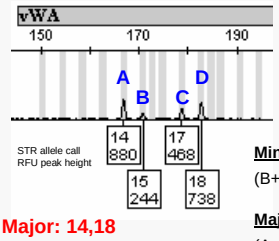
Minor component:
(A+D)/total = (438+523)/3397 = **0.283**

Major component:
(B+C)/total = (1110+1326)/3397 = **0.717**

Major: 23,24
Minor: 19,25

Four Peaks **(4 allele loci)**
- heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

Mix Ratio
4 Allele Locus: vWA



Total of all peak heights
= 880 + 244 + 468 + 738
= 2330 RFUs

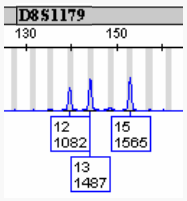
Minor component:
(B+C)/total = (244+468)/2330 = **0.306**

Major component:
(A+D)/total = (880+738)/2330 = **0.694**

Major: 14,18
Minor: 15,17

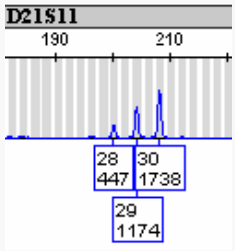
Four Peaks **(4 allele loci)**
- heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

3 Allele Locus: D8S1179



Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

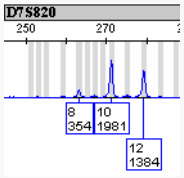
3 Allele Locus: D21S11



Allele	Frequency
28	0.159
29	0.195
30	0.278

Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

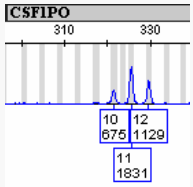
3 Allele Locus: D7S820



Allele	Frequency
8	0.151
10	0.243
12	0.166

Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

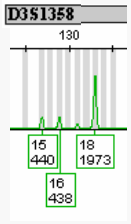
3 Allele Locus: CSF1PO



Allele	Frequency
10	0.217
11	0.301
12	0.361

Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

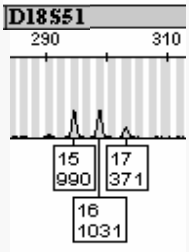
3 Allele Locus: D3S1358



Allele	Frequency
15	0.262
16	0.253
18	0.152

Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

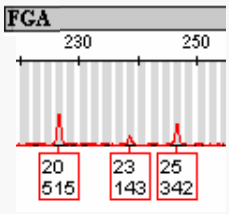
3 Allele Locus: D18S51



Allele	Frequency
15	0.159
16	0.139
17	0.126

Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

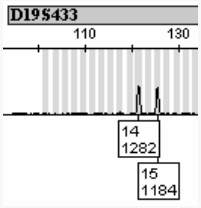
3 Allele Locus: FGA



Allele	Frequency
20	0.127
23	0.134
25	0.071

Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)

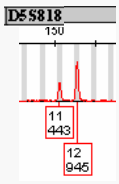
2 Allele Locus: D19S433



Allele	Frequency
14	0.369
15	0.152

Two Peaks **(2 allele loci)**
- heterozygote + heterozygote, two overlapping alleles (genotypes are identical)
- heterozygote + homozygote, one overlapping allele
- homozygote + homozygote, no overlapping alleles (genotypes are unique)

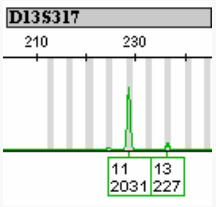
2 Allele Locus: D5S818



Allele	Frequency
11	0.361
12	0.384

Two Peaks **(2 allele loci)**
- heterozygote + heterozygote, two overlapping alleles (genotypes are identical)
- heterozygote + homozygote, one overlapping allele
- homozygote + homozygote, no overlapping alleles (genotypes are unique)

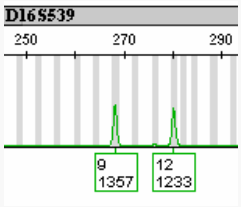
2 Allele Locus: D13S317



Allele	Frequency
11	0.339
13	0.124

Two Peaks **(2 allele loci)**
- heterozygote + heterozygote, two overlapping alleles (genotypes are identical)
- heterozygote + homozygote, one overlapping allele
- homozygote + homozygote, no overlapping alleles (genotypes are unique)

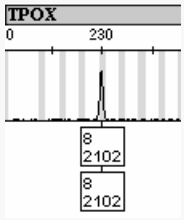
2 Allele Locus: D16S539



Allele	Frequency
9	0.113
12	0.326

Two Peaks **(2 allele loci)**
- heterozygote + heterozygote, two overlapping alleles (genotypes are identical)
- heterozygote + homozygote, one overlapping allele
- homozygote + homozygote, no overlapping alleles (genotypes are unique)

1 Allele Locus: TPOX



Allele	Frequency
8	0.535

Single Peak **(1 allele loci)**
- homozygote + homozygote, overlapping allele (genotypes are identical)

Profiles Used In Mixture Samples

	Victim	Suspect
D8S1179	13,15	12,12
D21S11	29,30	28,30
D7S820	10,12	8,10
CSF1PO	11,12	10,11
D3S1358	18,18	15,16
TH01	7,9	8,9.3
D13S317	11,11	11,13
D16S539	9,12	9,12
D2S1338	23,24	19,25
D19S433	14,15	14,15
vWA	14,18	15,17
TPOX	8,8	8,8
D18S51	15,16	16,17
AMEL	X,X	X,Y
D5S818	12,12	11,11
FGA	20,25	20,23

Software Programs (~~Expert Systems~~) for Mixture Deconvolution

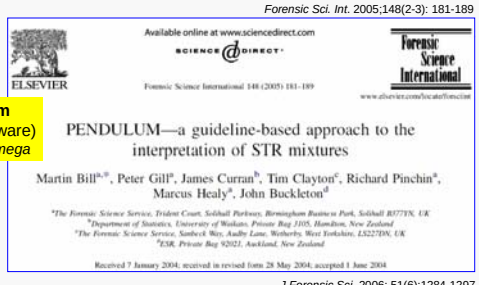
These programs do not supply stats (only attempt to deduce mixture components)

- Linear Mixture Analysis (LMA) **U.S. Patent 6,807,490**
  - **Part of TrueAllele system** developed by Mark Perlin (Cybergenetics)
  - Perlin, M. W. and Szabady, B. (2001) Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *J.Forensic Sci.* 46(6): 1372-1378
- Least Squares Deconvolution (LSD)
  - Available for use at **https://lsd.lit.net/**
  - Wang, T., Xue, N., Birdwell, J.D. (2006) Least-square deconvolution: a framework for interpreting short tandem repeat mixtures. *J Forensic Sci.* 51(6):1284-1297.
- PENDULUM
  - **Part of FSS i-3 software suite (i-STReam)**
  - Bill, M., Gill, P., Curran, J., Clayton, T., Pinchin, R., Healy, M., and Buckleton, J. (2005) PENDULUM-a guideline-based approach to the interpretation of STR mixtures. *Forensic Sci.Int.* 148(2-3): 181-189

USACIL program developed by Tom Overson called **DNA_DataAnalysis**

Forensic Sci. Int. 2005;148(2-3): 181-189

i-STReam (FSS-i³ software) *Sold by Promega*


Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Forensic Science International

PENDULUM—a guideline-based approach to the interpretation of STR mixtures

Martin Bill[a,*], Peter Gill[a], James Curran[b], Tim Clayton[c], Richard Pinchin[a], Marcus Healy[a], John Buckleton[d]

[a]The Forensic Science Service, Trident Court, Solihull Parkway, Birmingham Business Park, Solihull B377YN, UK
[b]Department of Statistics, University of Waikato, Private Bag 3105, Hamilton, New Zealand
[c]The Forensic Science Service, Sanbeck Way, Audby Lane, Wetherby, West Yorkshire, LS227DN, UK
[d]ESR, Private Bag 92021, Auckland, New Zealand

Received 7 January 2004; received in revised form 28 May 2004; accepted 1 June 2004

J Forensic Sci. 2006; 51(6):1284-1297

Available for use over internet at **https://lsd.lit.net/**

J Forensic Sci, November 2006, Vol. 51, No. 6
doi:10.1111/j.1556-4029.2006.00268.x
Available online at: www.blackwell-synergy.com

Tsewei Wang,[1] Ph.D.; Ning Xue,[1] M.Sc.; and J. Douglas Birdwell,[2] Ph.D.

Least-Square Deconvolution: A Framework for Interpreting Short Tandem Repeat Mixtures*

Acknowledgments

- Amy Christen (Marshall University NEST Project Team)
- Angie Dolph (NIST intern/Marshall University)
- Tim Kalafut (USACIL)

Mixture Statistics

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

German Type A,B, and C mixture classifications

- **Type A**, where major/minor contributors cannot be deduced, require stats
  - LR
  - RMNE
- **Type B** enables major contributor to be deduced
  - RMP (which is 1/LR)
- **Type C** no stats should be attempted because of the possibility of failure to account for allele dropout due to stochastic effects with low level DNA samples

Statistical Interpretation of DNA Mixtures


CROATIAN MEDICAL JOURNAL 42(3):244-246, 2001

CMJ FORENSIC SCIENCES

Interpretation of Complex Forensic DNA Mixtures

Carll Ladd, Henry C. Lee, Nicholas Yang, Frederick R. Bieber[1]

[1]State of Connecticut Department of Public Safety, Division of Scientific Services, Forensic Science Laboratory, Conn; and [2]Departments of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Mass, USA

Ladd et al. 2001. *Croatian Medical Journal* 43(3): 244-246

1. **Qualitative statement ('..cannot exclude..')**
2. **Interpret as single source from peak height differences, differential extraction, etc. and calculate random match probability (RMP)**
3. **Calculate probability of exclusion (CPE)**
4. **Calculate likelihood ratio (LR)**

Random Man Not Excluded (RMNE)

- = Probability of Exclusion (PE)
- John Buckleton (*Forensic DNA Evidence Interpretation*, p. 222) quotes Laszlo Szabo of Tasmania Forensic Science Laboratory: "Intuitively, RMNE is easier to explain to a jury and express in reports than the likelihood ratio, and is probably closer to what the court wants—e.g., the suspect matches the mixture, but what if this is the wrong person– then what is the probability that someone else in the population would also match the mixture (i.e., not be excluded as a contributor)."
- Buckleton (*Forensic DNA Evidence Interpretation*, p. 222) also quotes Bruce Weir: that exclusion probabilities "often rob the items of probative value"

Probability of Exclusion (RMNE)

- **Advantages**
  - Does not require an assumption of the number of contributors to a mixture
  - Easier to explain in court
- **Disadvantages**
  - Weaker use of the available information (robs the evidence of its true probative power because this approach does not consider the suspect's genotype)
  - Likelihood ratio approaches are developed within a consistent logical framework

John Buckleton, *Forensic DNA Evidence Interpretation*, p. 223

RMNE (CPE)

- Statements from DAB Recommendations on Statistics (FDT2e, p. 617)
- CPE provides a calculation of the estimated proportion of individuals from a defined population group **that can be excluded as a contributor** to an observed DNA mixture

Probability of Exclusion

The probability that a random person (unrelated individual) would be excluded as a contributor to the observed DNA mixture

For each locus, 1 minus the square of the sum of frequencies for the observed alleles

$$PE_i = 1 - \left(\sum_{i=1}^{n} p(A_i) \right)^2$$

Buckleton (2005) *Forensic DNA Evidence Interpretation*, p. 219

Across multiple loci (i.e., **combined probability of exclusion**, CPE):

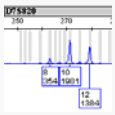
$$PE = 1 - \prod_{l} (1 - PE_l)$$

Buckleton (2005) *Forensic DNA Evidence Interpretation*, p. 221

Combined Probability of Exclusion (CPE)

Each locus is calculated separately and then combined for CPE

$CPE = 1 - (1 - PE_1)(1 - PE_2)(1 - PE_3)\ldots(1\text{-}PE_N)$



Probability of exclusion at a single locus:

- ***The combined frequency of alleles detected (P)***
  P = frequency of allele 1 + frequency of allele 2 + frequency of allele 3, … N

US Caucasian Data

Allele	Frequency
8	0.151
10	0.243
12	0.166

P = 0.151 + 0.243 + 0.166 = **0.56**

- ***The combined frequency of alleles not detected (Q)***
  Q = 1 – P

Q = 1 – 0.56 = **0.44**

- **PE = $Q^2 + 2Q(1-Q)$**

CPI = 1 – CPE

PE = $(0.44)^2 + 2(0.44)(1\text{-}0.44)$ = 0.1936 + 0.4928
PE = 0.686

Calculation from CPI Perspective

Each locus is calculated separately and then combined for CPE

CPI or $P_{profile}$ = (P_{locus1}) (P_{locus2}) (P_{locus3}) … ($P_{locus(N)}$)

Probability of inclusion at a single locus:

- ***Individual frequencies are summed and then squared***
  PI or $P_{locus} = (p_1 + p_2 + p_3 + \ldots + p_N)^2$
  Alleles present in the mixture; Remaining possible alleles in the population

- PE = 1 – P_{locus} = 1 – PI
- **PE = $Q^2 + 2Q(1-Q)$**

Essentially $P^2 + 2PQ + Q^2 = 1$
PI PE

P + Q = 1 so
P = 1 – Q and
Q = 1 – P

Provides probability of an unrelated individual in the population is a contributor to the mixture at the loci examined

Likelihood Ratios

Basic Math Terms

- When '+' is used, this means 'OR'
- When 'x' is used, this means 'AND'
- Pr. is shorthand for probability

- Therefore…
  - the probability of *a* 'AND' *b* happening together is **Pr(*a* and *b*) = *a* x *b***
  - the probability of *a* 'OR' *b* happening together is **Pr(*a* or *b*) = *a* + *b***

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Conditioning

- **Probabilities are conditional**, which means that the probability of something is based on a hypothesis

- In math terms, conditioning is denoted by a vertical bar
  - Hence, **Pr(a|b)** means 'the probability of *a* ***given*** that *b* is true"

- The probability of an event *a* is dependent upon various assumptions—and these assumptions or hypotheses can change…

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (1)

Defining the Event and Assumptions/Hypotheses

- Let's suppose that *a* is the probability of an event (e.g., **will it rain?**)
- What is the probability that it will rain in the afternoon – **Pr(a)**?
- This probability is dependent upon assumptions
  - We can look at the window in the morning and observe if it is sunny (s) or cloudy (c)
  - Pr(a) **if** it is sunny (s) is less than Pr(a) **if** it is cloudy (c)
- We can write this as **Pr(a|s)** and **Pr(a|c)**
  - Since sunny or cloudy are the only possibilities, Pr(s) + Pr(c) = 1
  - or **Pr(s) = 1 – Pr(c)**

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (2)

Examining Available Data

- Pr(a|s) and Pr(a|c) can be calculated from data
- How often does it rain in the afternoon when its sunny in the morning?
  - 20 out of 100 observations so **Pr(a|s) = 0.2**
- How often does it rain in the afternoon when it is cloudy in the morning?
  - 80 out of 100 observations so **Pr(a|c) = 0.8**

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (3)

Formation of the Likelihood Ratio (LR)

- The LR compares two probabilities to find out which of the two probabilities is the most likely

The probability that it will rain in the afternoon when it is cloudy in the morning or **Pr(a|c)** is divided by the probability that it will rain in the afternoon when it is sunny in the morning or **Pr(a|s)**

$$LR = \frac{\Pr(a \mid c)}{\Pr(a \mid s)} = \frac{0.8}{0.2} = 4$$

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (4)

Explanation of the Likelihood Ratio

$$LR = \frac{\Pr(a \mid c)}{\Pr(a \mid s)} = \frac{0.8}{0.2} = 4$$

- The probability that it will rain is 4 times more likely **if** it is cloudy in the morning than **if** it is sunny in the morning.
- The word **if** is very important here. It must always be used when explaining a likelihood ratio otherwise the explanation could be misleading.

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Likelihood Ratios in Forensic DNA Work

- We evaluate the evidence (*E*) relative to alternative pairs of hypotheses
- Usually these hypotheses are formulated as follows:
  - The probability of the evidence if the crime stain originated with the suspect or Pr(*E*|*S*)
  - The probability of the evidence if the crime stain originated from an unknown, unrelated individual or Pr(*E*|*U*)

$$LR = \frac{\Pr(E \mid S)}{\Pr(E \mid U)}$$

The numerator

The denominator

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

The Likelihood Ratio Must Be Stated Carefully

- The probability of the evidence is *x* times more likely **if** the stain came from the suspect Mr. Smith than **if** it came from an unknown, unrelated individual.
- It is not appropriate to say: "The probability that the stain came from Mr. Smith." because we must always include the conditioning statement – i.e., **always make the hypothesis clear in the statement.**
- Always use the word '**if**' when using a likelihood ratio to avoid this trap

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Likelihood Ratio (LR)

- Provides ability to express and evaluate both the prosecution hypothesis, H_p (the suspect is the perpetrator) and the defense hypothesis, H_d (an unknown individual with a matching profile is the perpetrator)

$$LR = \frac{H_p}{H_d}$$

- **The numerator, H_p, is usually 1** – since in theory the prosecution would only prosecute the suspect if they are 100% certain he/she is the perpetrator
- The denominator, **H_d**, is typically the profile frequency in a particular population (based on individual allele frequencies and assuming HWE) – i.e., **the random match probability**

Relationship between Likelihood Ratio (LR) and Random Match Probability (RMP)

- For single source samples or deduced major component profiles in a mixture…

$$LR = \frac{1}{RMP} \quad \text{or} \quad RMP = \frac{1}{LR}$$

Example #1

A Single Locus from a 2-Person Mixture

- Consider a simple **two person mixture** with one locus consisting of two heterozygotes with non-overlapping alleles
- If the suspect is *ab*, then there must be another (unknown person) who is *cd*

Suspect = a,b

Forget peak heights for the time being

Adapted from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Example #1

The Two Hypotheses Are Formed…

- **Prosecution (H_p)**: The DNA result has come from the suspect and one unknown person, or **Pr(*E*|*S*,*U*)**
- **Defense (H_d)**: The DNA result has come from two unknown people, or **Pr(*E*|U_1,U_2)**

$$LR = \frac{\Pr(E \mid S,U)}{\Pr(E \mid U_1,U_2)}$$

Suspect = a,b

Forget peak heights for the time being

Adapted from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Example #1

Formulating the Numerator
(Prosecution Hypothesis)

- If the prosecution hypothesis is true, then we would expect genotype *ab* to be present with 100% probability or Pr=1.
- The chance of seeing an unknown person of type *cd* is the frequency of that type in the population or $2p_cp_d$, where p_c is the allele frequency for allele c.
- **Pr(*E*|*S*,*U*)** = 1 x $2p_cp_d$ **= $2p_cp_d$**

Suspect = a,b

Forget peak heights for the time being

Adapted from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Example #1

Formulating the Denominator
(Defense Hypothesis)

- The defense claims that the evidence could come from any two random individuals
- We must work out **all possible pairwise combinations** from alleles *abcd* and their probabilities (genotype frequencies)

Multiplied because you are considering individual #1 **AND** individual #2

Added because you are considering any of the possibilities (combination 1,2,3,4,5, **OR** 6)

Individual #1	Individual #2	Products
ab	cd	$2p_ap_b \times 2p_cp_d$ → $4p_ap_bp_cp_d$
ac	bd	$2p_ap_c \times 2p_bp_d$ → $4p_ap_bp_cp_d$
ad	bc	$2p_ap_d \times 2p_bp_c$ → $4p_ap_bp_cp_d$
cd	ab	$2p_cp_d \times 2p_ap_b$ → $4p_ap_bp_cp_d$
bd	ac	$2p_bp_d \times 2p_ap_c$ → $4p_ap_bp_cp_d$
bc	ad	$2p_bp_c \times 2p_ap_d$ → $4p_ap_bp_cp_d$
Sum of products		**$24p_ap_bp_cp_d$**

Reverse combinations

Pr(*E*|U_1,U_2) = $24p_ap_bp_cp_d$

Adapted from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Example #1

Formulating the Likelihood Ratio

- The numerator and denominator are combined to form the LR
- And common elements in both numerator and denominator are eliminated to simplify the algebraic equation…

$$LR = \frac{\Pr(E \mid S,U)}{\Pr(E \mid U_1,U_2)} = \frac{\cancel{2}\,p_c p_d}{\cancel{24}^{12}\,p_a p_b p_c p_d} = \frac{1}{12\,p_a p_b}$$

All LR Calculations Follow the Same Basic Rules Just Shown

- Form hypotheses
  - Keep in mind what you are conditioning on
- The LR numerator belongs to the prosecution
- The LR denominator belongs to the defense
- Numerator and denominator are combined and equation is simplified
- Allele frequency values are placed into the equation for each locus
- **The LR from each locus is combined through multiplication if the loci are independently inherited (i.e., the product rule) to form a LR for the entire profile**

Example #2

Another Example…

- The evidentiary mixture profile is from a semen stained vaginal swab and possesses alleles a, b, c, and d.
- The suspect is a,b and the victim is c,d.
- Because it is reasonable to assume that the victim's alleles would be present on the swab (i.e., an intimate sample), we can condition on this…

Suspect = a,b
Victim = c,d

Example #2

With an Intimate Sample, the Hypothesis Changes…

- **Prosecution (H_p)**: The DNA result has come from the suspect and the victim, or **Pr(*E*|*S,V*)**
- **Defense (H_d)**: The DNA result has come from the victim and one unknown person, or **Pr(*E*|*U,V*)**

$$LR = \frac{\Pr(E \mid S,V)}{\Pr(E \mid U,V)}$$

Suspect = a,b
Victim = c,d

Example #2

Formulating the Numerator (Prosecution Hypothesis)

- The prosecution hypothesis (S+V) is completely explains the evidence. Hence, the probability is Pr=1
- **Pr(*E*|*S,V*) = 1 x 1 = 1**

Suspect = a,b
Victim = c,d

Example #2

Formulating the Denominator (Defense Hypothesis)

- The defense hypothesis is that the presence of alleles *a* and *b* are the result of an unknown person – and they concede that alleles *c* and *d* come from the victim
- Since the frequency of an unknown, unrelated individual possessing alleles *a* and *b* in the population is $2p_ap_b$, where p_a is the allele frequency for allele *a* and p_b is the allele frequency for allele *b*, then
- **Pr(*E*|*U,V*)** = $2p_ap_b$ x 1 = **$2p_ap_b$**

Suspect = a,b
Victim = c,d

Example #2

Formulating the Likelihood Ratio

- The numerator and denominator are combined to form the LR

$$LR = \frac{\Pr(E \mid S,V)}{\Pr(E \mid U,V)} = \frac{1}{2p_a p_b}$$

- **Note that this LR is the same as for a non-mixed sample comprising the suspect alone.**
- This example then is an illustration of simplification by "subtraction" (victim's alleles are being removed from mathematical consideration…).

Adapted from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Forming the Denominator (H_d) for the LR…

Evidence (Mixture)	Victim	Suspect	LR
A_1, A_2, A_3	A_2, A_3	A_1, A_2	$\frac{1}{p_1(2p_2+2p_3+p_1)}$
8,10,12	10,12	8,10	

Potential Combinations:

If victim is A_2,A_3, then perpetrator could be

	Type	Frequency (probability)
H_d	A_1,A_2	$2p_1p_2$
	A_1,A_3	$2p_1p_3$
	A_1,A_1	p_1^2

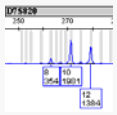
Other possible genotypes contributing to the evidence

Determine joint probability through summing individual probabilities

$2p_1p_2 + 2p_1p_3 + p_1^2 \longrightarrow p_1(2p_2 + 2p_3 + p_1)$

Likelihood Ratio (LR) Calculations

Evidence (Mixture)	Victim	Suspect	LR
A_1, A_2, A_3	A_2, A_3	A_1, A_2	$\frac{1}{p_1(2p_2+2p_3+p_1)}$
8,10,12	10,12	8,10	



US Caucasian Data

Allele	Frequency
A_1 8	p_1 0.151
A_2 10	p_2 0.243
A_3 12	p_3 0.166

$$LR = \frac{1}{(0.151)[(2)(0.243)+2(0.166)+(0.151)]}$$

LR = 6.83

Does not consider peak height information

The prosecution hypothesis (that the suspect is the perpetrator) is 6.83 times more likely than the defense hypothesis (that an unknown, unrelated individual is the perpetrator).

Likelihood Ratios for the Following Hypotheses

H_p: The mixture contains the DNA of the victim and the suspect

H_d: The mixture contains the DNA of the victim and an unknown, unrelated individual

Evidence (Mixture)	Victim	Suspect	LR
A_1, A_2, A_3, A_4	A_1, A_2	A_3, A_4	$\frac{1}{2p_3p_4}$
A_1, A_2, A_3	A_1, A_2	A_1, A_3 or A_2, A_3 or A_3, A_3	$\frac{1}{p_3(2p_1+2p_2+p_3)}$
A_1, A_2, A_3	A_1, A_1	A_2, A_3	$\frac{1}{2p_2p_3}$
A_1, A_2	A_1, A_2	A_1, A_1 or A_1, A_2 or A_2, A_2	$\frac{1}{(p_1+p_2)^2}$
A_1, A_2	A_1, A_1	A_1, A_2 or A_2, A_2	$\frac{1}{p_2(2p_1+p_2)}$
A_1, A_1	A_1, A_1	A_1, A_1	$\frac{1}{p_1^2}$

Adapted from Buckleton (2005) *Forensic DNA Evidence Interpretation*, Table 7.1, p. 229

DAB Recommendations on Statistics

February 23, 2000

Forensic Sci. Comm. 2(3); available on-line at http://www.fbi.gov/hq/lab/fsc/backissu/july2000/dnastat.htm

"The DAB finds either one or both PE or LR calculations acceptable and strongly recommends that one or both calculations be carried out whenever feasible and a mixture is indicated"

- Probability of exclusion (PE)
  - Devlin, B. (1993) Forensic inference from genetic markers. *Statistical Methods in Medical Research*, 2, 241–262.
- Likelihood ratios (LR)
  - Evett, I. W. and Weir, B. S. (1998) *Interpreting DNA Evidence*. Sinauer, Sunderland, Massachusetts.

Topics and Techniques for Forensic DNA Analysis

Interlaboratory Mixture Studies

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

Outline

- Purpose of Interlaboratory Studies
- Overview of Mixture Studies and Lessons Learned
- NIST MIX05 Study Results

Interlaboratory Studies

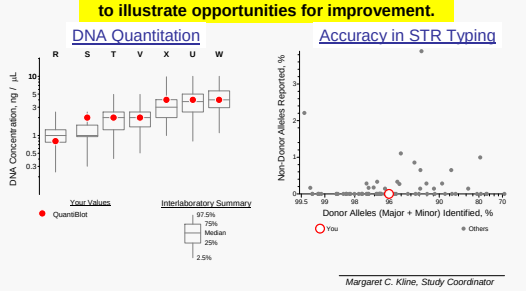
- Purpose…
  - Not a proficiency test
  - Most labs see them as opportunity to anonymously directly compare themselves to others
- STRBase section on interlab studies
  - http://www.cstl.nist.gov/biotech/strbase/interlab.htm

A High Degree of Variability Currently Exists with Mixture Interpretation

- **"If you show 10 colleagues a mixture, you will probably end up with 10 different answers"**
  - Peter Gill, Human Identification E-Symposium, April 14, 2005
- **Interlaboratory studies help to better understand why variability may exist between laboratories**
- Most analysts are only concerned about their own lab protocols and do not get an opportunity to see the big picture from the entire community that can be provided by a well-run interlaboratory study

The National Institute of Standards and Technology
Gratefully Acknowledges the Participation of the

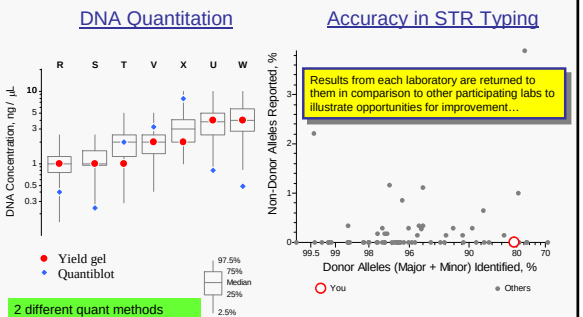
This feedback can be helpful to a laboratory to know where they stand relative to other labs to illustrate opportunities for improvement.



Margaret C. Kline, Study Coordinator

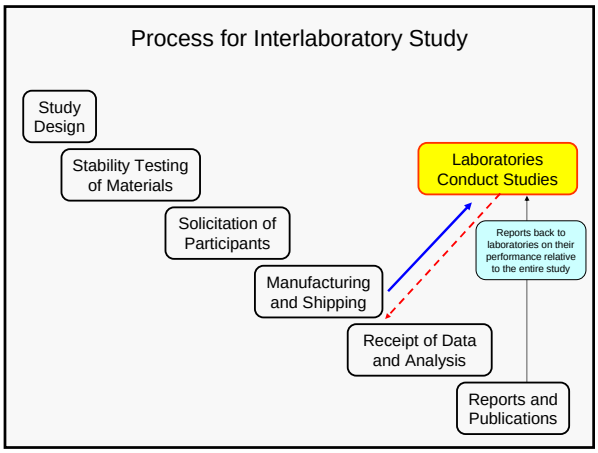
See Kline, M.C., *et al.* (2003) *Anal. Chem.* 75: 2463-2469

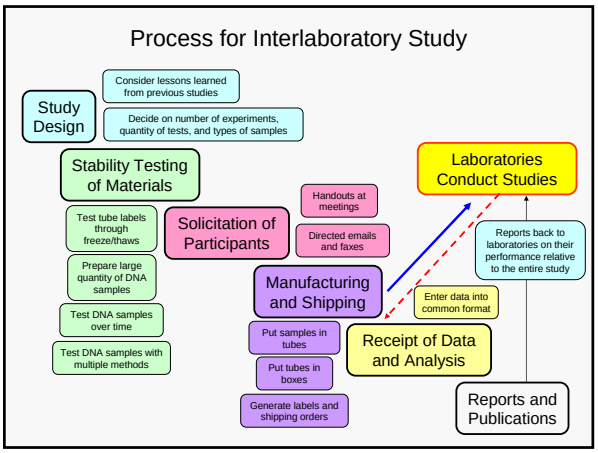
Individual Performance in an Interlaboratory Study



2 different quant methods gave different results; this lab followed the Quantiblot results

See Kline, M.C., *et al.* (2003) *Anal. Chem.* 75: 2463-2469





NIST Initiated Interlaboratory Studies

Studies involving STRs	# Labs	Publications
Evaluation of CSF1PO, TPOX, and TH01	34	Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M. (1997) Interlaboratory evaluation of STR triplex CTT. *J. Forensic Sci.* 42: 897-906
Mixed Stain Studies #1 and #2 (Apr–Nov 1997 and Jan–May 1999)	45	Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ. (2001) NIST Mixed Stain Studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *J. Forensic Sci.* 46: 1199-1210
MSS3 **Mixed Stain Study** #3 (Oct 2000-May 2001)	74	Kline, M.C., Duewer, D.L., Redman, J.W., Butler, J.M. (2003) NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Anal. Chem.* 75: 2463-2469. Duewer, D.L., Kline, M.C., Redman, J.W., Butler, J.M. (2004) NIST Mixed Stain Study #3: signal intensity balance in commercial short tandem repeat multiplexes, *Anal. Chem.* 76: 6928-6934.
DNA Quantitation Study (Jan-Mar 2004) QS04	80	Kline, M.C., Duewer, D.L., Redman, J.W., Butler, J.M. (2005) Results from the NIST 2004 DNA Quantitation Study, *J. Forensic Sci.* 50(3):571-578
Mixture Interpretation Study (Jan - Aug 2005) **MIX05**	**69**	***Several presentations made ...*** **Poster at 2005 Promega meeting (Sept 2005); available on STRBase**

Overall Lessons Learned from NIST MSS 1,2,&3

- Laboratories have instruments with different sensitivities
- **Different levels of experience and training plays a part in effective mixture interpretation**
- Amount of input DNA makes a difference in the ability to detect the minor component (labs that put in "too much" DNA actually detected minor components more frequently)

NIST MIX05 Summary

Purpose of MIX05 Study

- **Goal is to understand the "lay of the land" regarding mixture analysis across the DNA typing community**
- One of the primary benefits we hope to gain from this study is **recommendations for a more uniform approach to mixture interpretation** and training tools to help educate the community

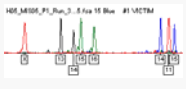
MIX05 Study Design and Purpose

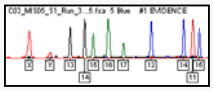
Interlab studies provide a "big picture" view of the community

- Permit a large number of forensic practioners to **evaluate the same mixture data**
- Provide multiple cases representing a range of mixture scenarios
- Generate data from multiple STR kits on the same mixture samples to compare performance for detecting minor components
- The primary variable should be the laboratory's interpretation guidelines rather than the DNA extraction, PCR amplification, and STR typing instrument sensitivity
- **Are there best practices in the field that can be advocated to others?**

Mixture Interpretation Interlab Study (MIX05)

- **Only involves interpretation of data – to remove instrument detection variability and quantitation accuracy issues**
- **94 labs enrolled** for participation
- **69 labs have returned results** (17 from outside U.S.)
- Four mock cases supplied with "victim" and "evidence" electropherograms (GeneScan .fsa files – that can be converted for Mac or GeneMapper; gel files made available to FMBIO labs)
- Data available with Profiler Plus, COfiler, SGM Plus, PowerPlex 16, Identifiler, PowerPlex 16 BIO (FMBIO) kits
- Summary of results will involve training materials to illustrate various approaches to solving mixtures





Perpetrator Profile(s) ??

Along with reasons for making calls and any stats that would be reported

Requests for Participants in MIX05

Mixtures representing four different case scenarios have been generated at NIST with multiple STR kits and provided to laboratories as electropherograms.

We would like to receive the following information:

1) **Report the results as though they were from a real case** including whether a statistical value would be attached to the results. *Please summarize the perpetrator(s) alleles in each "case" as they might be presented in court—along with an appropriate statistic (if warranted by your laboratory standard operating procedure) and the source of the allele frequencies used to make the calculation.* Please indicate which kit(s) were used to solve each case.

2) **Estimate the ratio for samples present in the evidence mixture** and how this estimate was determined.

3) **Provide a copy of your laboratory mixture interpretation guidelines** and a brief explanation as to why conclusions were reached in each scenario

A MIX05 Participant Noted…

"Things we do not do:

- **Calculate mixture ratios for casework**
  - **Calculation used for this study:** Find loci with 4 alleles (2 sets of sister alleles). Make sure sister alleles fall within 70%, then take the ratio of one allele from one sister set to one allele of the second sister set, figure ratios for all combinations and average. Use peak heights to calculate ratios.
- **Provide allele calls in reports**
- **Provide perpetrator(s) alleles or statistics in court without a reference sample to compare to the DNA profile obtained from the evidence. We will try to determine the perpetrator(s) profile for entry into CODIS."**

We recognize that some of the information requested in this interlab study may not be part of a lab's standard operating procedure

MIX05 Case Scenarios

Genomic DNA samples with specific allele combinations ("evidence") were mixed in the following ratios:

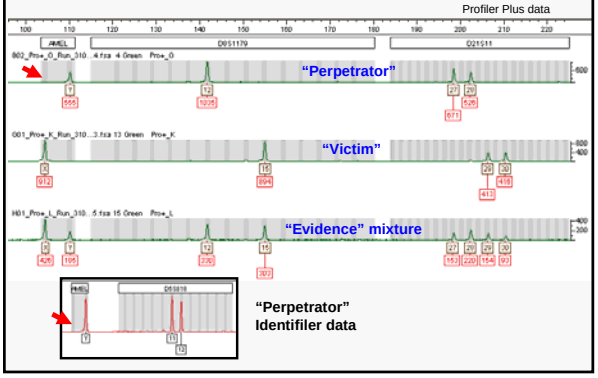
Based on Identifiler 15 STR loci

	#alleles		#loci with #alleles				
	N all	N unq	N 1	N 2	N 3	N 4	N 5
Case #1 – victim is major contributor (3F:1M)	39	26	2	6	5	2	0
Case #2 – perpetrator is major contributor (1F:3M)	55	52	0	1	4	10	0
Case #3 – balanced mixture (1F:1M) • Male lacked amelogenin X	48	37	0	3	8	4	0
Case #4 – more extreme mixture (7F:1M) • Male contained tri-allelic pattern at TPOX	50	42	0	3	7	4	1

Female victim DNA profile was supplied for each case

Labs asked to deduce the perpetrator DNA profile – suspect(s) not provided

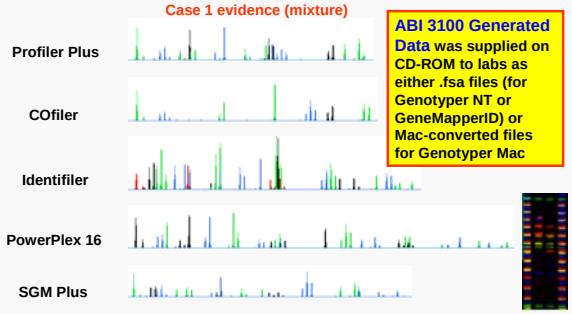
Amelogenin X allele is missing in male perpetrator DNA sample for MIX05 Case #3



MIX05 Results on Multiple Kits

http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm

Case 1 evidence (mixture)



ABI 3100 Generated Data was supplied on CD-ROM to labs as either .fsa files (for Genotyper NT or GeneMapperID) or Mac-converted files for Genotyper Mac

FMBIO data was also made available upon request

Summary of MIX05 Responses

94 labs enrolled for participation

69 labs returned results (17 from outside U.S.)

50 labs made allele calls
39 labs estimated ratios
29 labs provided stats

All participants were supplied with all data and could choose what kits to examine based on their experience and lab protocols

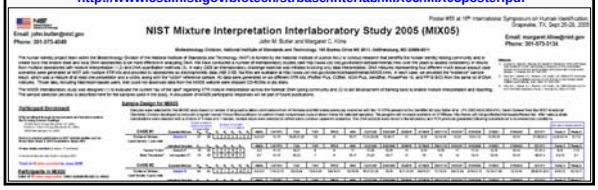
STR kit results used
- 34 ProfilerPlus/COfiler
- 10 PowerPlex 16
- 7 PP16 BIO
- 5 Identifiler
- 2 SGM Plus
- 1 All ABI kit data
- 9 Various combinations

Generally Identifiler data was of poorer quality in the electropherograms we provided...which caused some labs to not return results (they indicated a desire for higher quality data through sample re-injection to reduce pull-up prior to data interpretation)

What MIX05 Participants Have Received Back from NIST…

- Certificate of participation in the interlab study
- Copy of the poster presented at the Promega Sept 2005 meeting displaying "correct" results for the perpetrator in each case scenario as well as an explanation of study design and preliminary results

http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05/MIX05poster.pdf


NIST Mixture Interpretation Interlaboratory Study 2005 (MIX05)

When is a Sample a Potential Mixture?

According to several MIX05 participant interpretation guidelines

- Number of Observed Peaks
  - Greater than two peaks at a locus
  - More than two alleles are present at two or more loci, although three banded patterns can occur
  - Presence of 3 alleles at a single locus within a profile
  - 4 peaked patterns (if observed at any locus), 3 peaked patterns (if observed at two or more loci), significant imbalances (peak height ratios <60%) of alleles for a heterozygous genotype at two or more loci with the exception of low template amplifications, which should be interpreted with caution
- Imbalance of heterozygote alleles
  - thresholds range from 50-70%
- Stutter above expected levels
  - generally 15-20%

Detection thresholds also varied in the range of 50-200 RFUs

These protocol differences can lead to variation in reported alleles and therefore the deduced profile and resulting statistics

Summary of Some MIX05 Reported Results

Case #2 has perpetrator as major component and thus is the easiest to solve...

CASE #2		D3S1358	VWA	FGA	AMEL	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820	D16S539	TH01	TPOX	CSF1PO
True Perp	Z779619	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
LabID	Kit Used														
16	ProPlus/Cofiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--
6	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
91	SGM Plus	15	15	20,24	X,Y	11,13	28,32.2	17,18				10,11	7,9.3		
46	PP16	--	--	--	--	--	--	--	--	--	--	--	--	--	--
37	ProPlus/Cofiler	--	15	20	X,Y	13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
2	PP16	15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	INC	8,10	10,11	7,9.3	9,10	7,10
13	PP16 & Identifiler	15	15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
34	ProPlus/Cofiler	15	15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
70	Identifiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
55	ProPlus/Cofiler	15	15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
21	ProPlus/Cofiler	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
73	ProPlus/Cofiler	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
29	Identifiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
54	All Kits	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
90	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
9	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
4	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
33	ProPlus/Cofiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--
12	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
67	PP16	15	15,16	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
86	ProPlus/Cofiler	15,15	15,15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
79	ProPlus/Cofiler	15,15	15,15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
77	Identifiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--
60	PP16	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
61	Identifiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Most calls were correct (when they were made)

Some Mixture Ratios Reported in MIX05

Many labs do not routinely report the estimated ratio of mixture components

LabID	Case1 (F:M)	Case2 (M:F)	Case3 (M:F)	Case4 (F:M)
13	2	5	<2	10
34	1.8--3.6	3.9--6.7	1.6--1.8	6.2--7.6
70				
55	68%:32%	85%:15%	64%:36%	
21				
73	2:1	6:1	2:1	not determined
29				
54	2:1	6:1	2:1	6:1
90	male23-39%	not determined	male64-71%	
9	3 or 4:1	4 or 5:1	1.4:1	~10:1
4	10:1	6:1	1:1	not determined
33	male60-78%	male80-90%	male58-71%	victim86%
12	male25%	male85%	male40-45%	unknown10%
67	1:2.3	6.4:1	2:1	1:6.8
86	2:1	6-6.5:1	1.6-2:1	4-4.5:1
79	~3:1 to ~2:1	~6:1 to ~4:1	~2:1*	a lot of victim
77				
60	2:1	5:1	2:1	10:1
61				

Some Reported Stats for MIX05 Case #1

Many of the 29 labs providing statistics used PopStats 5.7

		Case1		
LabID	Kits Used	Caucasians	African Americans	Hispanics
77	Identifiler	PE calculated	PE calculated	PE calculated
73	ProPlus/Cofiler	none provided	none provided	none provided
4	ProPlus/Cofiler	none provided	none provided	none provided
12	ProPlus/Cofiler	none provided	none provided	none provided
29	Identifiler	none provided	none provided	none provided
90	ProPlus/Cofiler	1.18E+15	2.13E+14	3.09E+15
34	ProPlus/Cofiler	2.40E+11	7.00E+09	9.80E+10
46	PP16	5.60E+09	3.80E+11	none provided
33	ProPlus/Cofiler	2.94E+08	1.12E+08	1.74E+09
6	ProPlus/Cofiler	40,000,000	3,500,000	280,000,000
9	ProPlus/Cofiler	1.14E+07	1.97E+07	1.54E+08
61	Identifiler	1.50E+06	260,000	2.40E+07
79	ProPlus/Cofiler	930,000	47,900	1,350,000
16	ProPlus/Cofiler	434,600	31,710	399,100

Which loci are included in each calculation?

Some Differences in Reporting Statistics

		Case1		
LabID	Kits Used	Caucasians	African Americans	Hispanics
90	ProPlus/Cofiler	1.18E+15	2.13E+14	3.09E+15
34	ProPlus/Cofiler	2.40E+11	7.00E+09	9.80E+10
33	ProPlus/Cofiler	2.94E+08	1.12E+08	1.74E+09
6	ProPlus/Cofiler	40,000,000	3,500,000	280,000,000
9	ProPlus/Cofiler	4.14E+07	1.97E+07	1.54E+08
79	ProPlus/Cofiler	930,000	47,900	1,350,000
16	ProPlus/Cofiler	434,600	31,710	399,100

~10 orders of magnitude difference (10^5 to 10^{15}) based on which alleles were deduced and reported

Remember that these labs are interpreting the same MIX05 electropherograms

Further Examination of These 7 Labs

		Case 1	ASCLD-LAB accredited?	Solved loci listed?
LabID	Kits Used	Caucasians		
90	ProPlus/Cofiler	1.18E+15	Yes	Yes
34	ProPlus/Cofiler	2.40E+11	Yes	Yes
33	ProPlus/Cofiler	2.94E+08	Yes	No
6	ProPlus/Cofiler	40,000,000	Yes	Yes
9	ProPlus/Cofiler	4.14E+07	No	No (CPE)
79	ProPlus/Cofiler	930,000	Yes	Yes
16	ProPlus/Cofiler	434,600	Yes	No

Possible Reasons for Variability in Reported Statistics:

- Different types of calculations (CPE vs RMP)
- Different loci included in calculations (due to different thresholds used)
- Different allele frequency population databases (most use PopStats)
- Use of victim (e.g., major component in Case 1) profile stats

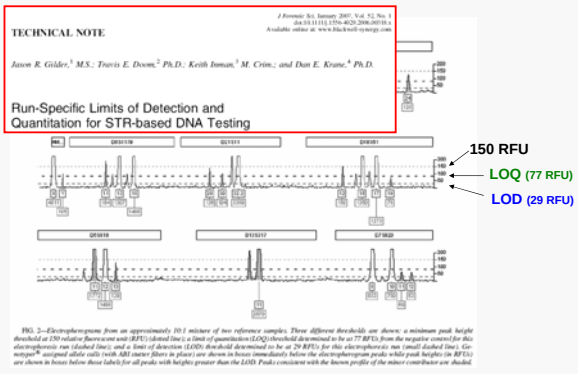
Different Stats Used

Combined Probability of Exclusion

Random Match Probability on Deduced Profiles

- **Lab 9** (4.14×10^7) used 1/CPI
- **Lab 6** (4.0×10^7) used selected loci and summed all possible genotypes for loci not completely deduced
- **Lab 90** (1.18×10^{15}) used theta value of 0.03 and deduced alleles at all 13 loci (***correctly deduced all perpetrator alleles***)

Different Thresholds of Detection Influence Allele Calls



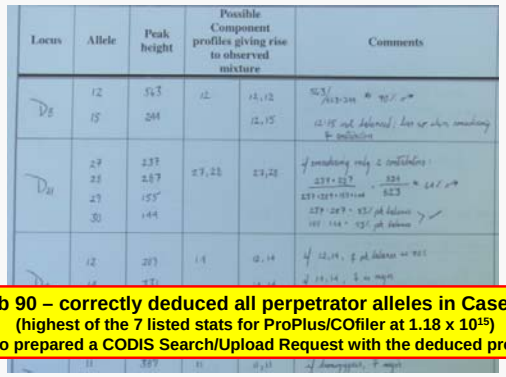
Gilder, J.R., Doom, T.E., Inman, K., Krane, D.E. (2007) Run-specific limits of detection and quantitation for STR-based DNA testing. *J. Forensic Sci.* 52(1): 97-101.

Different Detection Thresholds Used

		Case 1	
LabID	Kits Used	Caucasians	
90	ProPlus/Cofiler	1.18E+15	**75 RFUs**; all 13 STRs; **all results correct**
34	ProPlus/Cofiler	2.40E+11	**Not stated**; 8 STRs, 2 partial, 3 INC
33	ProPlus/Cofiler	2.94E+08	**75 RFUs**; no deduced alleles reported
6	ProPlus/Cofiler	40,000,000	**Not provided**; 3 STRs, 6 partial, 4 INC
9	ProPlus/Cofiler	4.14E+07	**100 RFUs**; no deduced alleles reported
79	ProPlus/Cofiler	930,000	**150 RFUs**; 2 STR, 5 partial, 6 INC
16	ProPlus/Cofiler	434,600	**Not stated**; no deduced alleles reported

- **Lab 90** has **specific, detailed mixture interpretation guidelines** with worked examples and a fabulous flowchart
- **Lab 16** has **vague guidelines** that begin with "mixture interpretation is not always straightforward. Analysts must depend on their knowledge and experience…"

Manually Solving Mixture Component Profiles



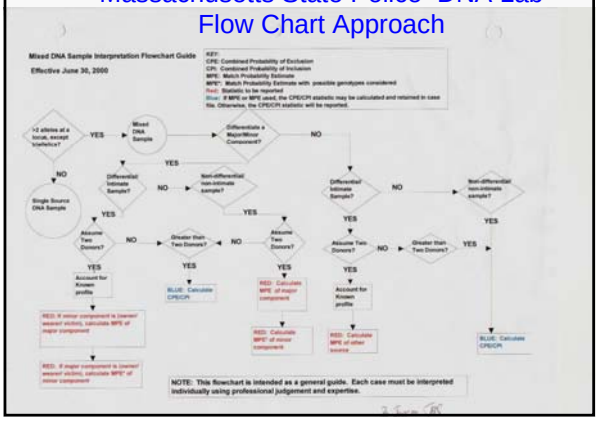
Lab 90 – correctly deduced all perpetrator alleles in Case #1
(highest of the 7 listed stats for ProPlus/COfiler at 1.18 x 10^{15})
Also prepared a CODIS Search/Upload Request with the deduced profile

A Model Report of Analysis…

- "The Profiler Plus and COfiler sample files were evaluated by **four different analysts**, using both NT and MAC analysis platforms. **The analysts checked for concordance, and a single conclusion for each mock case has been issued**."
- They detailed all assumptions made outside the course of routine casework:
  - Assumed intimate samples
  - That a comparison of deduced "foreign" alleles had been made with the perpetrator's known standard in order to calculate the significance of the inclusion with the evidentiary profile
- For Case #4: "A **Combined Probability of Inclusion was calculated** and reported for only those loci where all the alleles were above threshold [75 RFUs]. However, a minor profile(s) could not be deduced from this sample. **Please note that our laboratory may employ strategies to gain more information from the sample, such as a 10 second injection of the CE and Y-STR analysis**.

Lab 90

Massachusetts State Police DNA Lab Flow Chart Approach



Quotes from One Lab's MIX05 Report

- Case 1: STR typing results from the Evidence sample indicate a DNA mixture profile. The victim cannot be excluded as a possible donor of the genetic material in the Evidence sample. No statistics will be generated at this time.
- The Evidence samples would have to be rerun in order to verify any alleles called in the final profiles. This is true for any mixed sample profiles as per our laboratory guidelines.
- **Our laboratory does not "pull out" any profile from a mixture for interpretation or statistical purposes.** The exception to this is for CODIS profiles where the alleles that can be unambiguously attributed to the victim are removed.
- **We currently do not calculate and report statistics on mixture samples.**

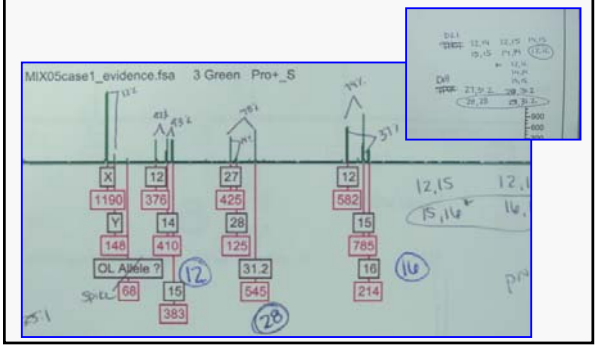
Lab 88

Examples of MIX05 Report Formats

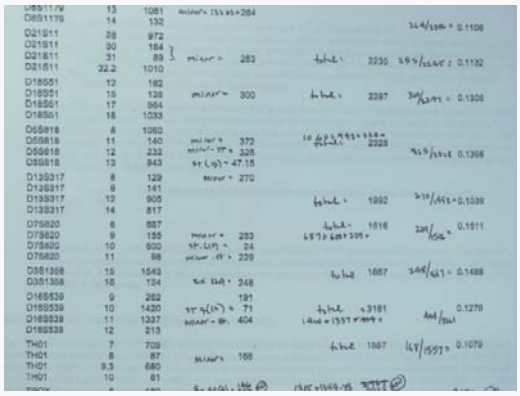
All examples with Case #1

(~3:1 mixture with female victim as the major component – and victim profile is provided)

Manual Solving of MIX05 Peak Ratios and Possible Mixture Combinations



Another MIX05 Participant Manually Solving a Mixture



Semi-Automated Locus-by-Locus Interpretation Performed by One MIX05 Participant

D21S11

Peak	A	B	C	D
Allele	28	30	31	32.2
RFU	988	167	92	1025

Known type: K1 30, K2 31

Possible combinations					Thresholds 70%	Thresholds 60%	Htrzyg. Pair 1	Htrzyg. Pair 2	Mixture ratio(1/2)	
28	30	AND	31	32.2	N	N	16.90%	8.98%		
28	31	AND	30	32.2	N	N	9.31%	16.29%		
28	32.2	AND	30	31	N	N	96.39%	55.09%		Best fit

D18S51

Peak	A	B	C	D
Allele	12	15	17	18
RFU	163	139	972	1047

Known type: K1 12, K2 15

Possible combinations					Thresholds 70%	Thresholds 60%	Htrzyg. Pair 1	Htrzyg. Pair 2	Mixture ratio(1/2)	
12	15	AND	17	18	Y	Y	85.28%	92.84%	0.15	Known present
12	17	AND	15	18	N	N	16.77%	13.28%		
12	18	AND	15	17	N	N	15.57%	14.30%		

Excel spreadsheet used to examine possible component combinations

Different Reporting Formats for MIX05 Data

Different Reporting Formats for MIX05 Data

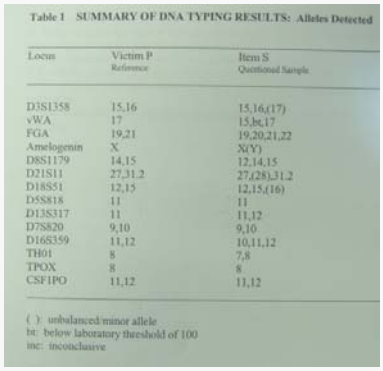


Table 1 SUMMARY OF DNA TYPING RESULTS: Alleles Detected

Locus	Victim P Reference	Item S Questioned Sample
D3S1358	15,16	15,16,(17)
vWA	17	15,bt,17
FGA	19,21	19,20,21,22
Amelogenin	X	X(Y)
D8S1179	14,15	12,14,15
D21S11	27,31.2	27,(28),31.2
D18S51	12,15	12,15,(16)
D5S818	11	11
D13S317	11	11,12
D7S820	9,10	9,10
D16S359	11,12	10,11,12
TH01	8	7,8
TPOX	8	8
CSF1PO	11,12	11,12

( ): unbalanced minor allele
bt: below laboratory threshold of 100
inc: inconclusive

No attempt to deduce perpetrator alleles (foreign profile)

Different Reporting Formats for MIX05 Data

Profile that would be put into CODIS

LOCI	CODIS ENTRY * obligate allele	OTHER ALLELE'S IN SUSPECT'S POSSIBLE PROFILE
D3S1358	17	16,17
VWA	15*	15,17
FGA	20,22	20,22
D8S1179	12	12,12
D21S11	28*	28,31.2
D18S51	15*	15,16
D5S818	--	
D13S317	12	12,12
D7S820	--	10
D16S539	10,11*	10,11
THO1	7*	7,8 maybe
TPOX	8	8 maybe
CSF1PO	--	11,12 maybe

Kits – Profiler Plus and Cofiler
Ratio - 1:2 (perpetrator:victim)

Different Reporting Formats for MIX05 Data

	Items	
Locus	"S" Case 1 Evid.	"P" Case 1 Victim
D3S1358	15, 16, *	15, 16
D16S539	(10), 11, (12)	11, 12
AMEL	X, *	X
THO1	(7), 8	8
TPOX	8	8
CSF1PO	11, 12	11, 12
D7S820	9, 10	9, 10
vWA	(15), 17	17
FGA	19, 20, 21, 22	19, 21
D8S1179	12, 14, 15	14, 15
D21S11	27, 31.2, *	27, 31.2
D18S51	12, 15, (16)	12, 15,
D5S818	11	11
D13S317	11, 12	11

() indicates apparent minor peaks in a mixture.
"*" indicates peaks below the VFL threshold (150 rfus) for reporting.

Different Reporting Formats for MIX05 Data

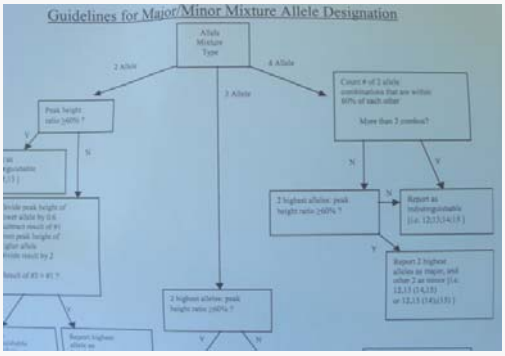
Case 1:

Item: description	D3S1358	VWA	FGA	AMEL	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820	D16S539	TH01	TPOX	CSF1PO
Pro+/CO_S: evid 1	15,16 (17)	15|17	19|20| 21|22	X,X (Y)	12|14|15	27,31.2 (28)	12,15 (16)	11,11	11|12	9|10	10|11 |12	7|8	8,8	11,12
Pro+/CO_P: victim 1 reference	15,16	17,17	19,21	X,X	14,15	27,31.2	12,15	11,11	11,11	9,10	11,12	8,8	8,8	11,12
Male interpreted from evidence 1	17	15,15 15,17	20,22	X,Y	12,12	28	16	11,11	12,12	Nd	10,11	7,7 7,8	Nd	Nd

Two allele values separated by a comma represent a genotype. Genotype calls assume biallelic donors with no null alleles.
() Indicates minor allele detected.
Single numbers and numbers separated by "|" represent an allele only designation rather than a genotype.
Interpreted profile assumes that the victim is present in the evidence mixture of two people. More than one genotype may be listed where a single genotype could not be conclusively determined. Nd=not determined due to level of results.

The community would benefit from more uniform reporting formats and mixture solving strategies…

Some Protocols Have Flow Charts to Help Make Decisions in Mixture Resolution



Value of the MIX05 Study

http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm

- Data sets exist with multiple mixture scenarios and a variety of STR kits that **can be used for training purposes**
- A wide variety of approaches to mixture interpretation have been applied on the **same data sets evaluated as part of a single study**
- **Interpretation guidelines from many laboratories are being compared to one another for the first time in an effort to determine challenges facing future efforts to develop "expert systems" for automated mixture interpretation**
- **We are exploring the challenges of supplying a common data set to a number of forensic laboratories** (e.g., if a standard reference data set was ever desired for evaluating expert systems)

Conclusions from the MIX05 Study (Opportunities for Improvement)

- It is worth taking a closer look at protocol differences between labs to see the impact on recovering information from mixture data
- Training should help bring greater consistency
- Expert systems (when they become available and are used) should help aid consistency in evaluating mixtures and help produce more uniform reporting formats

NIST Software Programs to Aid Mixture Work

Excel-based programs developed by David Duewer (NIST)

- **mixSTR** *(developed at request of Palm Beach Sheriff's Office)*
  - Does not interpret data (relies on user inputted alleles following STR data review)
  - Aids in the organization of STR mixture information
  - Considers only the presence/absence of alleles (no peak heights used)
- **Virtual MixtureMaker** *(developed to aid MIX05 sample selection)*
  - Creates mixture combinations through pairwise comparisons of input STR profiles
  - Returns information on the number of loci possessing 0,1,2,3,4,5, or 6 alleles in each 2-person mixture (also reports number of loci in each sample with 0,1,2, or 3 alleles)
  - Useful for selection of samples in mixture or validation studies with various degrees of overlapping alleles in combined STR profiles
  - Useful in checking for potentially related individuals in a population database

Programs can be downloaded from NIST STRBase web site:
http://www.cstl.nist.gov/div831/strbase/software.htm

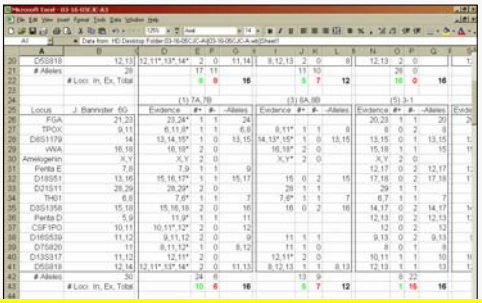
mixSTR Program

Comparisons are made between

- suspect and evidence (S/E) alleles,
- suspect and suspect (S/S) alleles **(to look for potential close relatives),**
- evidence and other evidence (E/E) sample(s) alleles **(to see how various evidentiary samples compare to one another),** and
- controls to evidence (C/E) and controls to suspect (C/S) alleles **(as a quality control contamination check).**

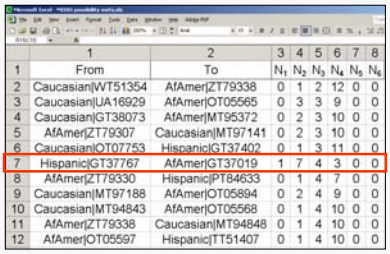
mixSTR S/E output

Data from Palm Beach County Sheriff's Office Case
Supplied by Catherine Cothran



Example of suspect to evidence (S/E) comparisons made in this case. Note that the suspect is 21,23 at FGA while the evidence contains 23,24* (* indicates that allele 24 is a minor component). Thus this suspect has allele 23 in common and is missing allele 24 in the evidence.

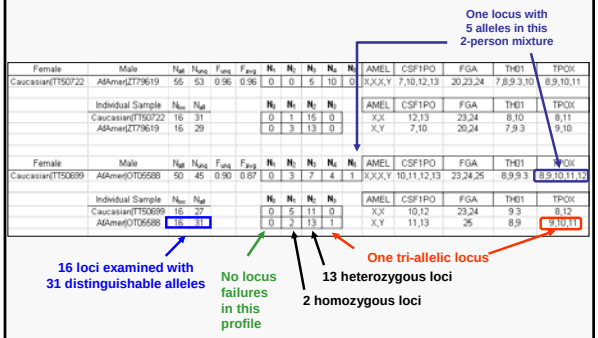
Virtual MixtureMaker Output



	1	2	3	4	5	6	7	8
1	From	To	N_1	N_2	N_3	N_4	N_5	N_6
2	Caucasian|WT51354	AfAmer|ZT79338	0	1	2	12	0	0
3	Caucasian|UA16929	AfAmer|OT05565	0	3	3	9	0	0
4	Caucasian|GT38073	AfAmer|MT95372	0	2	3	10	0	0
5	AfAmer|ZT79307	Caucasian|MT97141	0	2	3	10	0	0
6	Caucasian|OT07753	Hispanic|GT37402	0	1	3	11	0	0
7	Hispanic|GT37767	AfAmer|GT37019	1	7	4	3	0	0
8	AfAmer|ZT79330	Hispanic|PT84633	0	1	4	7	0	0
9	Caucasian|MT97188	AfAmer|OT05894	0	2	4	9	0	0
10	Caucasian|MT94843	AfAmer|OT05568	0	1	4	10	0	0
11	AfAmer|ZT79338	Caucasian|MT94848	0	1	4	10	0	0
12	AfAmer|OT05597	Hispanic|TT51407	0	1	4	10	0	0

When the STR profiles for these two individuals are combined to create a 2-person mixture, the mixture profile will contain 1 locus with a single allele, 7 loci with two alleles, 4 loci with three alleles, and 3 loci with four alleles (and no loci with 5 or 6 alleles, which is only possible if one or both samples possess tri-allelic patterns at the same STR locus).

Virtual MixtureMaker Output



Some Final Thoughts…

- It is of the highest importance in the art of detection to be able to recognize out of a number of facts, which are incidental and which vital. Otherwise your energy and attention must be dissipated instead of being concentrated (Sherlock Holmes, *The Reigate Puzzle*).
- "**Don't do mixture interpretation unless you have to**" (Peter Gill, Forensic Science Service, 1998).
- Mixture interpretation consumes a large part of DNA analysts' time – software tools that improve consistency in analysis will speed casework reporting and hopefully cases solved

Conclusion

ELSEVIER — Forensic Science International

DNA mixtures in forensic casework: a 4-year retrospective study

Yolanda Torres[a,*], Inmaculada Flores[a], Victoria Prieto[a], Manuel López-Soto[a], María José Farfán[a], Angel Carracedo[b], Pilar Sanz[a]

[a]Instituto Nacional de Toxicología, A. Pestal 863, E-41080 Sevilla, Spain
[b]Institute de Medicina Legal, Universidad de Santiago de Compostela, Sevilla, Spain

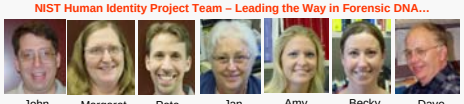
Received 27 March 2003; accepted 1 April 2003

"Mixture interpretation theory is well established and used in forensic laboratories. Most mixtures detected in casework are satisfactorily solved. But from this revision we can conclude that the behaviour of each mixed sample can be different and multifactorial and occasionally its interpretation turns out to be complicated—sometimes paralleling the importance of the evidence in the resolution of the case. In some casework mixtures our experience has proved that theoretical assumptions from studies with laboratory samples, albeit very useful, can turn out to be impracticable. **We consider that more sharing of day to day forensic laboratory problems is needed to refine our technical procedures in the resolution of specially difficult evidence."**

Acknowledgments

Funding from interagency agreement 2003-IJ-R-029 between NIJ and the NIST Office of Law Enforcement Standards

NIST Human Identity Project Team – Leading the Way in Forensic DNA…



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

Role in MIX05

- Margaret Kline (running study, sample prep, data review)
- John Butler (study design and data review)
- Becky Hill (GeneMapper*ID* data review)
- Jan Redman (Access database entry, shipping)
- Dave Duewer (*Virtual MixtureMaker* to aid sample selection; **mixSTR program**)
- Chris Tomsey & Frank Krist (FMBIO Mac data)
- Kermit Channel & Mary Robnett (FMBIO NT data)

Mandy Sozer for early discussions on study design

The many forensic scientists and their supervisors who took time out of their busy schedules to examine the MIX05 data provided as part of this interlaboratory study

Relevant Literature on Mixture Interpretation

General Information

Gill, P. (2002) Role of short tandem repeat DNA in forensic casework in the UK--past, present, and future perspectives. *BioTechniques* 32(2): 366-385.

Gill, P., Brenner, C.H., Buckleton, J.S., Carracedo, A., Krawczak, M., Mayr, W.R., Morling, N., Prinz, M., Schneider, P.M., Weir, B.S. (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101.

Gill, P., et al. (2008) National recommendations of the technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *FSI Genetics* 2: 76-82.

Ladd, C., Lee, H.C., Yang, N., Bieber, F.R. (2001) Interpretation of complex forensic DNA mixtures. *Croatian Med. J.* 42(3): 244-246.

Schneider, P.M., Gill, P., Carracedo, A. (2006) Editorial on the recommendations of the DNA commission of the ISFG on the interpretation of mixtures. *Forensic Sci. Int.* 160: 89.

Schneider, P.M., Fimmers, R., Keil, W., Molsberger, G., Patzelt, D., Pflug, W., Rothamel, T., Schmitter, H., Schneider, H., Brinkman, B. (2006) General recommendations of the (German) stain commission on the interpretation of DNA results from mixed stains. *Rechtsmedizin* 16:401-404. **(article in German)**

Torres, Y., Flores, I., Prieto, V., Lopez-Soto, M., Farfan, M.J., Carracedo, A., Sanz, P. (2003) DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic Sci. Int.* 134: 180-186.

Mixture Detection and Component Profile Deconvolution

Clayton, T.M., Whitaker, J.P., Sparkes, R., Gill, P. (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.* 91: 55-70.

Cowell, R.G., Lauritzen, S.L., Mortera, J. (2007) Identification and separation of DNA mixtures using peak area information. *Forensic Sci. Int.* 166(1):28-34

Egeland, T., Dalen, I., Mostad, P.F. (2003) Estimating the number of contributors to a DNA profile. *Int. J. Legal Med.* 117: 271-275.

Evett, I.W., Buffery, C., Willott, G., Stoney, D. (1991) A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *J. Forensic Sci. Soc.* 31: 41-47.

Evett, I.W., Gill, P.D., Lambert, J.A. (1998) Taking account of peak areas when interpreting mixed DNA profiles. *J. Forensic Sci*. 43(1): 62-69.

Evett, I.W., Foreman, L.A., Lambert, J.A., Emes, A. (1998) Using a tree diagram to interpret a mixed DNA profile. *J. Forensic Sci.* 43(3): 472-476.

Gill, P., Sparkes, R.L., Pinchin, R., Clayton, T.M., Whitaker, J.P., Buckleton, J.S. (1998) Interpreting simple STR mixtures using allelic peak areas. *Forensic Sci. Int.* 91: 41-53.

Shrestha, S., Strathdee, S.A., Broman, K.W., Smith, M.W. (2006) Unknown biological mixtures evaluation using STR analytical quantification. *Electrophoresis* 27: 409-415.

Tomsey CS, Kurtz M, Flowers B, Fumea J, Giles B, Kucherer S. (2001) Case work guidelines and interpretation of short tandem repeat complex mixture analysis. *Croatian Med. J.* 42: 276-280.

Designating True Alleles versus Artifacts

Gill, P., Sparkes, R., Kimpton, C. (1997) Development of guidelines to designate alleles using an STR multiplex system. *Forensic Sci. Int.* 89: 185-197.

Gill, P., Sparkes, R.L., Buckleton, J.S. (1998) Interpretation of simple mixtures when artifacts such as stutters are present—with special reference to multiplex STRs used by the Forensic Science Service. *Forensic Sci. Int.* 95: 213-224.

Kloosterman, A.D. and Kersbergen, P. (2003) Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci. *J. Soc. Biol.* 197(4): 351-359.

Leclair, B., Fregeau, C.J., Bowen, K.L., Fourney, R.M. (2004) Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples. *J. Forensic Sci.* 49: 968-980.

Leclair, B., Sgueglia, J.B., Wojtowicz, P.C., Juston, A.C., Fregeau, C.J., Fourney, R.M. (2003) STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. *J. Forensic Sci.* 48(5): 1001-1013.

Meldgaard, M. and Morling, N. (1997) Detection and quantitative characterization of artificial extra peaks following polymerase chain reaction amplification of 14 short tandem repeat systems used in forensic investigations. *Electrophoresis* 18: 1928-1935.

Sparkes, R., Kimpton, C., Watson, S., Oldroyd, N., Clayton, T., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A., Gill, P. (1996) The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies. *Int. J. Legal Med.* 109: 186-194.

Sparkes, R., Kimpton, C., Gilbard, S., Carne, P., Andersen, J., Oldroyd, N., Thomas, D., Urquhart, A., Gill, P. (1996) The validation of a 7-locus multiplex STR test for use in forensic casework. (II), Artefacts, casework studies and success rates. *Int. J. Legal Med.* 109: 195-204.

van Oorschot, R.A., Gutowski, S.J., Robinson, S.L., Hedley, J.A., Andrew, I.R. (1996) HUMTH01 validation studies: effect of substrate, environment, and mixtures. *J. Forensic Sci*. 41: 142-145.

Expert System Software Approaches

Bill, M., Gill, P., Curran, J., Clayton, T., Pinchin, R., Healy, M., and Buckleton, J. (2005) PENDULUM-a guideline-based approach to the interpretation of STR mixtures. *Forensic Sci. Int*. 148: 181-189.

Mortera, J., Dawid, A.P., Lauritzen, S.L. (2003) Probabilistic expert system for DNA mixture profiling. *Theor. Popul. Biol.* 63: 191-205.

Perlin, M. W. and Szabady, B. (2001) Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *J. Forensic Sci.* 46(6): 1372-1378.

Perlin, M.W. (2006) Scientific validation of mixture interpretation methods. *Proceedings of Promega's Seventeenth International Symposium on Human Identification*. Available at http://www.promega.com/geneticidproc/ussymp17proc/oralpresentations/Perlin.pdf

Wang, T., Xue, N., Birdwell, J.D. (2006) Least-squares deconvolution: a framework for interpreting short tandem repeat mixtures. *J. Forensic Sci.* 51(6): 1284-1297.

Interlaboratory Studies on Mixture Interpretation

Duewer, D.L., Kline, M.C., Redman, J.W., Newall, P.J., Reeder, D.J. (2001) NIST mixed stain studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *J. Forensic Sci.* 46(5): 1199-1210.

Duewer, D.L., Kline, M.C., Redman, J.W., Butler, J.M. (2004) NIST mixed stain study 3: signal intensity balance in commercial short tandem repeat multiplexes. *Anal. Chem.* 76: 6928-6934.

Kline, M.C., Duewer, D.L., Redman, J.W., Butler, J.M. (2003) NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Anal. Chem.* 75: 2463-2469.

Statistical Calculations and Issues

Balding, D.J. (2005) *Weight-of-evidence for Forensic DNA Profiles*. John Wiley & Sons; see mixture section on pp. 101-110.

Brenner, C.H., Fimmers, R., Baur, M.P. (1996) Likelihood ratios for mixed stains when the number of donors cannot be agreed. *Int. J. Legal Med.* 109:218-219.

Buckleton, J.S., Evett, I.W., Weir, B.S. (1998) Setting bounds for the likelihood ratio when multiple hypotheses are postulated. *Sci. Justice.* 38: 23-26.

Buckleton, J.S., Curran, J.M., Gill, P. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28.

Clayton, T. and Buckleton, J. (2005) Mixtures. Chapter 7 in *Forensic DNA Evidence Interpretation* (Eds.: Buckleton, J., Triggs, C.M., Walsh, S.J.), CRC Press, pp. 217-274.

Curran, J.M., Triggs, C.M., Buckleton, J.S., Weir, B.S. (1999) Interpreting DNA mixtures in structured populations. *J. Forensic Sci.* 44(5): 987-995.

Curran, J.M., Gill, P., Bill, M.R. (2005) Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure. *Forensic Sci. Int.* 148: 47-53.

Fung, W.K. and Hu, Y.-Q. (2001) The evaluation of mixed stains from different ethnic origins: general result and common cases. *Int. J. Legal Med.* 115: 48-53.

Fung, W.K. and Hu, Y.-Q. (2002) Evaluating mixed stains with contributors of different ethnic groups under the NRC-II Recommendation 4.1. *Stat. Med.* 21: 3583-3593.

Fung, W.K. and Hu, Y.-Q. (2002) The statistical evaluation of DNA mixtures with contributors from different ethnic groups. *Int. J. Legal Med.* 116: 79-86.

Fung, W.K., Hu, Y.Q., Chung, Y.K. (2006) On statistical analysis of forensic DNA: theory, methods and computer programs. Forensic Sci. Int. 162: 17-23.

Hu, Y.-Q. and Fung, W.K. (2003) Interpreting DNA mixtures with the presence of relatives. *Int. J. Legal Med*. 117: 39-45.

Hu, Y.-Q. and Fung, W.K. (2003) Evaluating forensic DNA mixtures with contributors of different structured ethnic origins: a computer software. Int J Legal Med. 117: 248-249.

Hu, Y.-Q. and Fung, W.K. (2005) Evaluation of DNA mixtures involving two pairs of relatives. *Int. J. Legal Med.* 119: 251-259.

Lauritzen, S.L. and Mortera, J. (2002) Bounding the number of contributors to mixed DNA stains. *Forensic Sci. Int.* 130:125-126.

Weir, B.S., Triggs, C.M., Starling, L., Stowell, L.I., Walsh, K.A.J., Buckleton, J.S. (1997) Interpreting DNA mixtures. *J. Forensic Sci.* 42(2): 213-222.

Defense Attacks on Mixture Interpretation

Gilder, J.R., Doom, T.E., Inman, K., Krane, D.E. (2007) Run-specific limits of detection and quantitation for STR-based DNA testing. *J. Forensic Sci.* 52(1): 97-101.

Paoletti, D.R., Doom, T.E., Krane, C.M., Raymer, M.L., Krane, D.E. (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. *J. Forensic Sci.* 50(6): 1361-1366.

Y-STRs Can Benefit Some Mixture Samples Compared to Autosomal STRs

Cerri N, Ricci U, Sani I, Verzeletti A, De Ferrari F. (2003) Mixed stains from sexual assault cases: autosomal or Y-chromosome short tandem repeats? *Croatian Med. J.* 44: 289-292.

Parson, W., Niederstatter, H., Brandstatter, A., Berger, B. (2003) Improved specificity of Y-STR typing in DNA mixture samples. *Int. J. Legal Med.* 117: 109-114.

Prinz M, Boll K, Baum H, Shaler B. (1997) Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci. Int.* 85: 209-218.

Y-STR Mixture and Statistical Issues

Butler, J.M., Decker, A.E., Kline, M.C., Vallone, P.M. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation. *J. Forensic Sci.* 50(4): 853-859.

Diederiche, M., Martin, P., Amorim, A., Corte-Real, F., Gusmao, L. (2005) A case of double alleles at three Y-STR loci: forensic implications. *Int. J. Legal Med.* 119: 223-225.

Fukshansky, N. and Bar, W. (2005) DNA mixtures: biostatistics for mixed stains with haplotypic genetic markers. *Int. J. Legal Med.* 119: 285-290.

Wolf, A., Caliebe, A., Junge, O., Krawczak, M. (2005) Forensic interpretation of Y-chromosomal DNA mixtures. *Forensic Sci Int.* 152: 209-213.