Agenda for Utah DNA Training
May 13-14, 2009
Salt Lake City, UT

John M. Butler, Ph.D.
National Institute of Standards and Technology

Wednesday, May 13, 2009

12:30 p.m. Background Information and Introductions

1:00 p.m. CE Fundamentals

2:30 p.m. *BREAK*

2:45 p.m. CE Troubleshooting

3:15 p.m. Y-STRs

4:30 p.m. *End first day* – provide homework assignment – mixture to solve

HOMEWORK

Thursday, May 14, 2009

8:00 a.m. qPCR and low level DNA testing

9:15 a.m. Mixture Interpretation

9:45 a.m. *BREAK*

10:00 a.m. Mixture Interpretation (cont.)

12:00 p.m. *Training concludes – awarding of certificates*

Reference Lists of Relevant Articles Supplied for Each Topic

Topics and Techniques for Forensic DNA Analysis
Continuing Education Seminar

Introductions

Utah DNA Training

Salt Lake City, UT
May 13-14, 2009

Dr. John M. Butler
National Institute of Standards and Technology
john.butler@nist.gov

Outline for Training Planned

May 13 (afternoon)
- Introductions
- Capillary Electrophoresis Fundamentals

BREAK

- CE Troubleshooting
- Y-STRs

HOMEWORK

May 14 (morning)
- qPCR and low level DNA testing
- Mixture Interpretation

BREAK

- Mixture Interpretation (cont.)

Dr. John M. Butler

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

B.S. Chemistry 1992

Experience

- University of Virginia/FBI Laboratory (1992-1995)
  - Work performed in Bruce McCord's lab
- NIST NRC Postdoc (1995-1997)
- GeneTrace Systems Inc (1997-1999)
- NIST Human Identity Project Leader (1999-present)

Contact Information
john.butler@nist.gov
301-975-4049

- Ph.D. dissertation (Aug 1995): "Sizing and quantitation of polymerase chain reaction products by capillary electrophoresis for use in DNA typing"
- *Forensic DNA Typing* textbook (now in its 2nd Edition)
- STRBase website: http://www.cstl.nist.gov/biotech/strbase/
- Family: wife Terilynne and 6 children
- Hobbies: reading, writing, and making PowerPoint slides

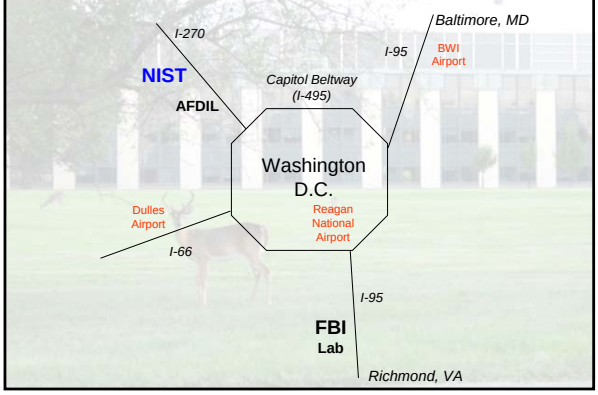
NIST History and Mission

- 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

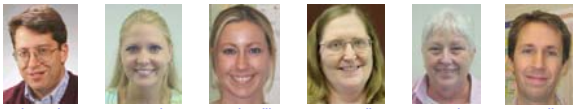
DNA typing standard

Location of NIST

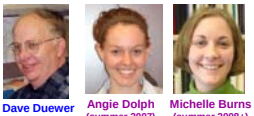


NIST Human Identity Project Team

...Bringing traceability and technology to the scales of justice...



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



Dave Duewer (data analysis) | **Angie Dolph** (summer 2007) | **Michelle Burns** (summer 2008+)

And many wonderful collaborators...

Since 2000:
>100 publications
>250 presentations
>30 training workshops

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

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

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

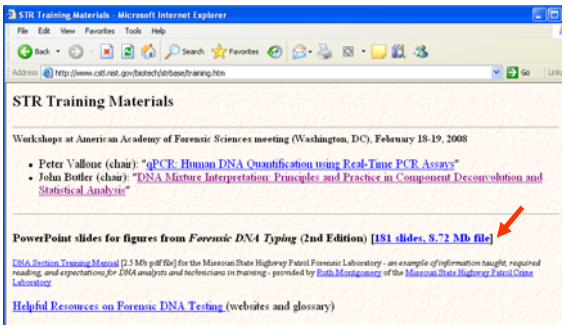
Current Areas of NIST Effort with Forensic DNA

- **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

Angie Dolph

Marshall U./ NIST

Amy Decker

NIST

mixtures

NIST and NIJ Disclaimer

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

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

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…

Background of Participants…

Your name
Your organization

What you hope to learn from this workshop

Topics and Techniques for Forensic DNA Analysis
Continuing Education Seminar

Capillary Electrophoresis Fundamentals

Utah DNA Training

Salt Lake City, UT
May 13-14, 2009

Dr. John M. Butler
National Institute of Standards and Technology
john.butler@nist.gov

Presentation Outline

- History and background on CE
- Separation
- Injection and sample preparation
- Detection
- Troubleshooting

My Goal:
To help you understand the basic chemistry behind DNA separations and to help make CE instruments less of a "black box"

Pioneers of Capillary Electrophoresis

Stellan Hjertén
Uppsala University

1967
First high voltage CE system (with rotating 3 mm i.d. capillaries)

James Jorgenson
University of North Carolina

1981
First "modern" CE experiments (with 75 µm i.d. capillaries)

Barry Karger
Northeastern University

1988/90
First DNA separations in a capillary (gel-filled/ sieving polymer)

Stellan Hjertén

Uppsala University (Sweden)

With first fully automated capillary free zone electrophoresis apparatus in 1967

In 2003 at age 75

Received his PhD (1967) under Professor **Arne Tiselius** who had developed moving boundary zone electrophoresis in 1937 **(Noble Prize in 1948)**

A Brief History of Capillary Electrophoresis

- 1937 – Tiselius develops moving boundary electrophoresis
- 1967 – Hjertén uses rotating 3 mm i.d. tubes for CE
- **1981 – Jorgenson and Lukacs demonstrate first high performance CE separations with 75 µm i.d. capillary**
- 1988 – Karger's group shows DNA separations of single stranded oligonucleotides with gel-filled capillaries
- 1990 – Karger's group shows DNA separations with sieving polymers on DNA restriction fragments
- 1991 – Grossman expands work with sieving polymers
- 1992 – Bruce McCord starts working on PCR product separations with STR allelic ladders

My Experience with CE, STRs, etc.

- May 1993 – began working in Bruce McCord's lab at Quantico
- Sept 1993 – developed mtDNA amplicon quantitation method (used in FBI casework from 1996 to present)
- Nov 1993 – **first demonstration of STR typing by CE** (using dual internal standards and TH01 ladder)
- July 1995 – defended Ph.D. dissertation entitled "Sizing and Quantitation of Polymerase Chain Reaction Products by Capillary Electrophoresis for Use in DNA Typing"
- July 1995 – ABI 310 Genetic Analyzer was released

My Experience with CE, STRs, etc. (cont.)

- 1996-1997 Developed STRBase while a postdoc at NIST
- Nov 1998 – GeneTrace Systems purchased a 310; typed several hundred samples with Profiler Plus and Cofiler kits and compared results to mass spec STR analysis
- **1999-present – Run thousands of samples with all STR kits available (except PP 1.2) and developed a number of new STR multiplex systems**
- Jan 2001 – Published "*Forensic DNA Typing: Biology and Technology behind STR Markers*" (2nd Edition in Feb 2005)
- April 2001-present – Use of ABI 3100 16-capillary array system

In the early 1990s the real question was how to transition from a gel to a capillary

- Cross-linked acrylamide gel filled capillaries were tried first
  - Reusable?
  - Bubble formation
  - Themal degradation
- Alternative was to not use a gel at all
  - Refillable sieving polymers
  - However, resolution was poor early on

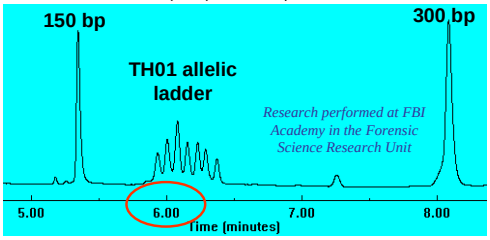
Early Work with CE and STRs

- Barry Karger's group (1988-1990)
  - Utilized gel-filled capillaries to separate ssDNA
  - Introduced sieving polymers in the form of linear polyacrylamide to separate restriction digests
- Beckman P/ACE 2050 is introduced in 1992 as the first commercially available CE coupled to a laser to enable fluorescence detection
- John Butler and Bruce McCord (1993-1995)
  - **First STR typing with single color CE using intercalating dyes and dual bracketing internal size standards**
- Rich Mathies' group (1995)
  - First STR typing with multi-color CE (and multi-capillary) using dye-labeled primers
- **ABI 310 is introduced in July 1995 as the first commercially available multi-color CE**

First Rapid STR Typing with Capillary Electrophoresis

Single color detection with dual internal size standards

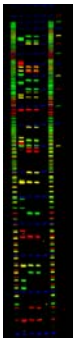
Butler *et al.* (1994) *BioTechniques* 17: 1062-1070



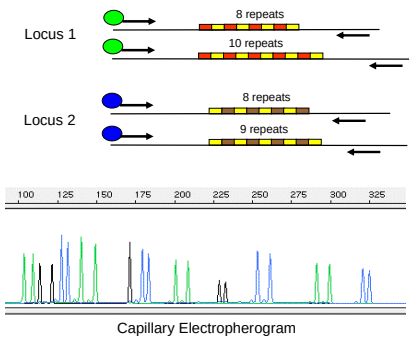
Performed in December 1993

Technology Implementation Takes Time – the FBI did not start running casework samples using STRs and CE until January 1999

STR Allele Separation Can Be Performed by Gel or Capillary Electrophoresis with Detection of Fluorescent Dyes Labeling Each PCR Product



Scanned **Gel Image**



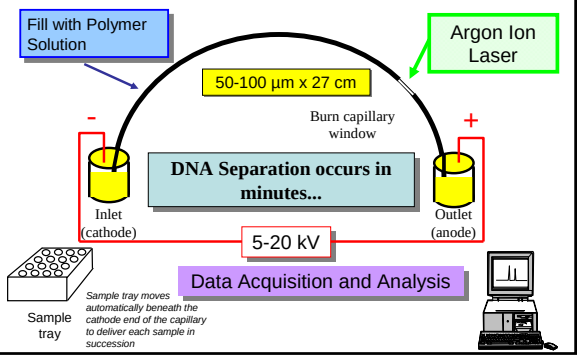
Capillary Electropherogram

Why Use CE for DNA Analysis?

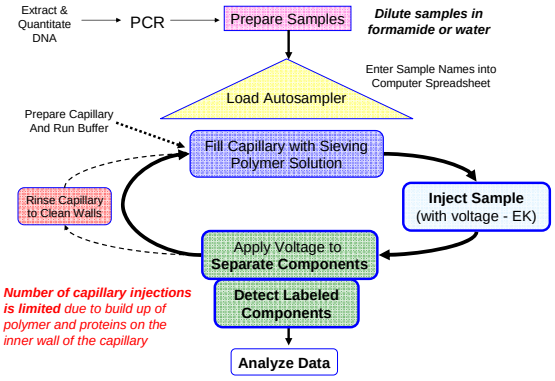
1. **Injection, separation, and detection are automated.**
2. **Rapid separations are possible**
3. **Excellent sensitivity and resolution**
4. **The time at which any band elutes is precisely determined**
5. **Peak information is automatically stored for easy retrieval**

Symbol first used in Oct 1994 at the Promega meeting when I had a poster introducing the use of CE for STR typing

Capillary Electrophoresis (CE)



Flowchart of CE Sample Analysis



Typical Instruments Used for STR Typing

GeneAmp 9700

Thermal Cycler for PCR Amplification

Capillary electrophoresis instruments for separating and sizing PCR products

single capillary

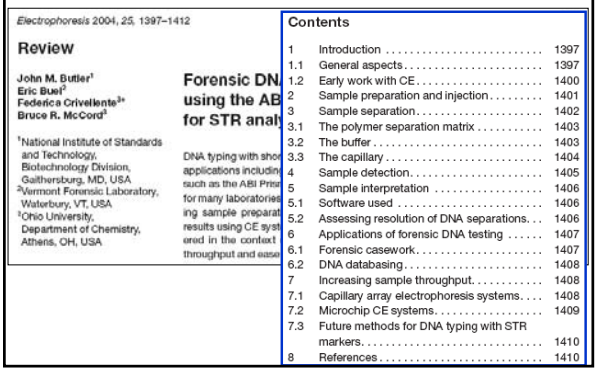
ABI 310

16-capillary array

ABI 3100

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

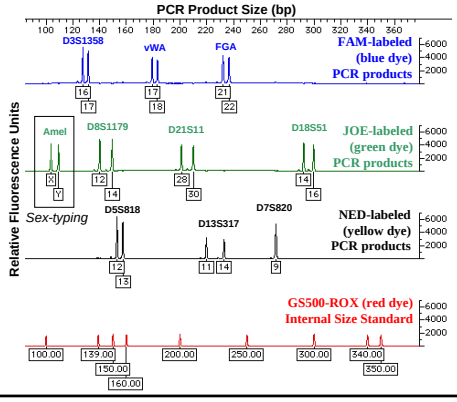
Forensic DN
using the AB
for STR anal

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

Contents

1	Introduction	1397
1.1	General aspects	1397
1.2	Early work with CE	1400
2	Sample preparation and injection	1401
3	Sample separation	1402
3.1	The polymer separation matrix	1403
3.2	The buffer	1403
3.3	The capillary	1404
4	Sample detection	1405
5	Sample interpretation	1406
5.1	Software used	1406
5.2	Assessing resolution of DNA separations	1406
6	Applications of forensic DNA testing	1407
6.1	Forensic casework	1407
6.2	DNA databasing	1408
7	Increasing sample throughput	1408
7.1	Capillary array electrophoresis systems	1408
7.2	Microchip CE systems	1409
7.3	Future methods for DNA typing with STR markers	1410
8	References	1410

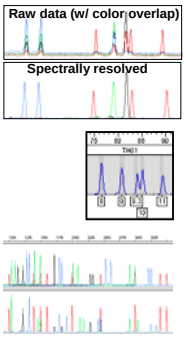
Genotype Results with Profiler Plus™ kit

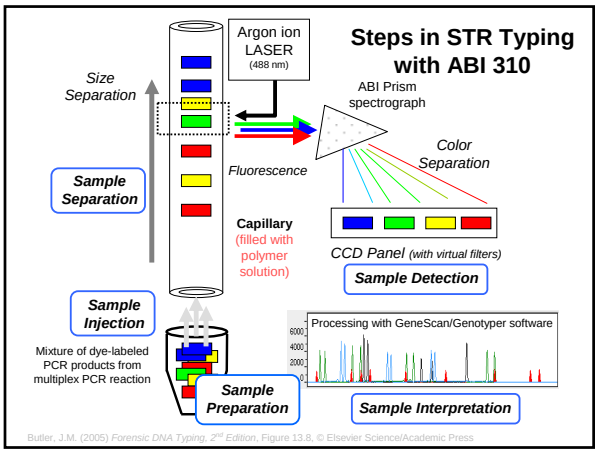


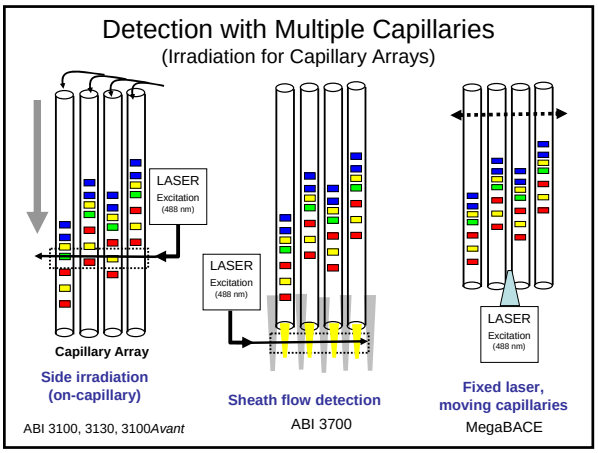
Analytical Requirements for STR Typing

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







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

Separation

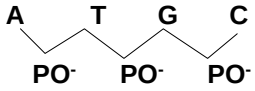
Ohm's Law

- 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)

DNA and Electrophoresis

"From a practical point of view it is disappointing that electrophoresis cannot be used to fractionate or analyze DNA's on the basis of size" Olivera, Biopolymers 1964, 2, 245

$\mu_{ep} = q/6\pi\eta r$ **small ions with high charge move fastest**



As size increases so does charge!

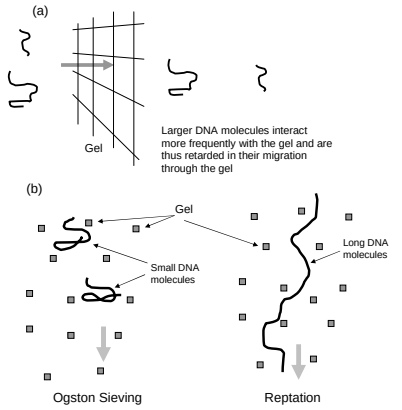


Figure 12.4, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Separation Issues

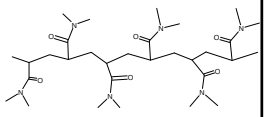
- Electrophoresis buffer –
  - Urea for denaturing and viscosity
  - Buffer for consistent pH
  - Pyrolidinone for denaturing DNA
  - EDTA for stability and chelating metals
- Polymer solution -- POP-4 (but others work also)
- Capillary wall coating -- dynamic coating with polymer
  - Wall charges are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

DNA Separations in Entangled Polymer Sieving Solutions

- Size based separation due to interaction of DNA molecules with entangled polymer strands
- Polymers are not cross-linked (as in slab gels)
- "Gel" is not attached to the capillary wall
- Pumpable -- can be replaced after each run
- Polymer length and concentration determine the separation characteristics

Entangled Polymer Solutions

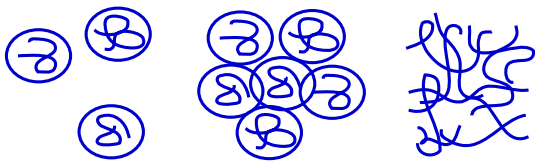
- Polymers are not cross-linked (above entanglement threshold)
- "Gel" is not attached to the capillary wall
- Pumpable -- can be replaced after each run
- Polymer length and concentration determine the separation characteristics
- Examples:
  - 1% HEC (hydroxyethyl cellulose)
  - 4% polyvinyl pyrolidinone
  - POP-4 and POP-6



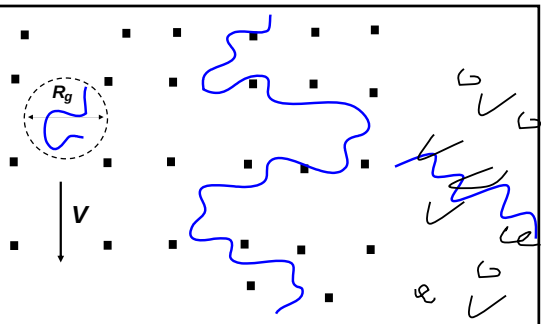
POP4 Polymer

Polydimethyl acrylamide

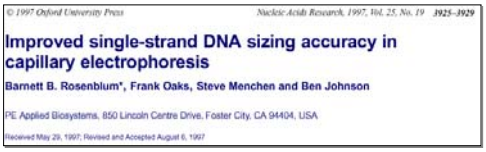
Transient Pores Are Formed Above the Entanglement Threshold.



C* < *C* ***C* = *C**** ***C* > *C****



What is in POP-4 and Genetic Analyzer Buffer?

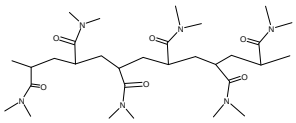


See also Wenz *et al.* (1998) *Genome Research* 8: 69-80

POP-4 (4% poly-dimethylacrylamide, 8 M urea, 5% 2-pyrrolidinone)

US Patent 5,552,028 covers POP-4 synthesis

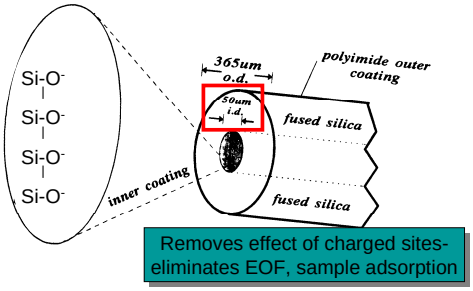
Running buffer contains 100 mM TAPS and 1 mM EDTA (adjusted to pH 8.0 with NaOH) *TAPS = N-Tris-(hydroxymethyl)methyl-3-aminopropane-sulfonic acid*



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 ºC); 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 ºC on the ABI 310, if Tris-borate was used, fluorescent intensity of PCR products would be lower

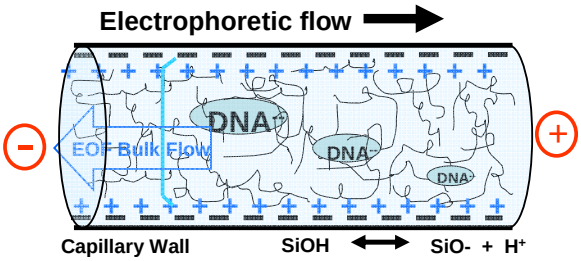
Capillary Coating



Removes effect of charged sites-eliminates EOF, sample adsorption

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

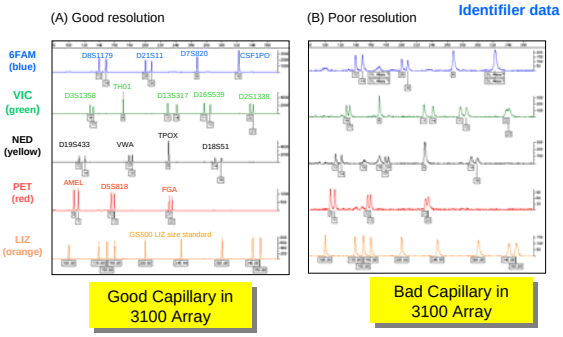
Capillary Wall Coatings Impact DNA Separations



Electroosmotic flow (EOF)

Solvated ions drag solution towards cathode in a flat flow profile

Capillary Resolution Differences



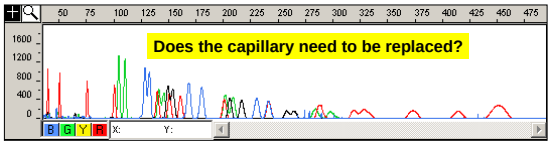
Good Capillary in 3100 Array

Bad Capillary in 3100 Array

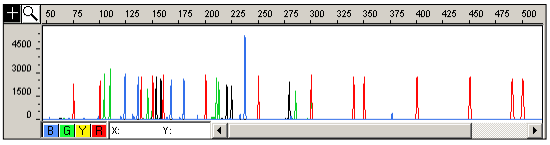
Butler, J.M., Buel, E., Crivellente, F., McCord, B.R. (2004) Forensic DNA typing by capillary electrophoresis: using the ABI Prism 310 and 3100 Genetic Analyzers for STR analysis. *Electrophoresis*, 25: 1397-1412.

What we call "melt downs"…

probably due to an incompletely filled capillary



No! The next injection looks fine…

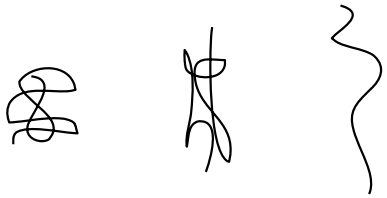


How to Improve Resolution?

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

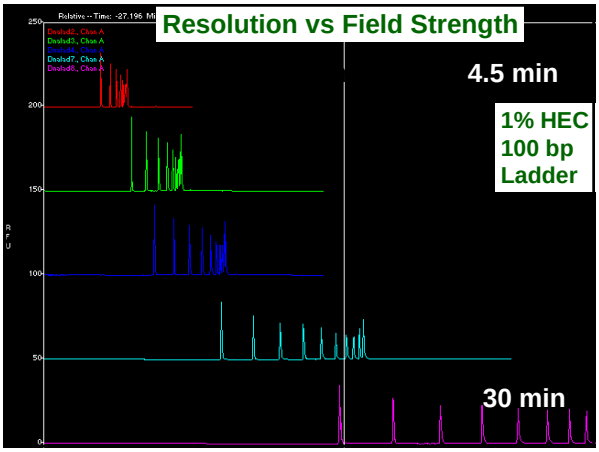
All of these come at a cost of longer separation run times

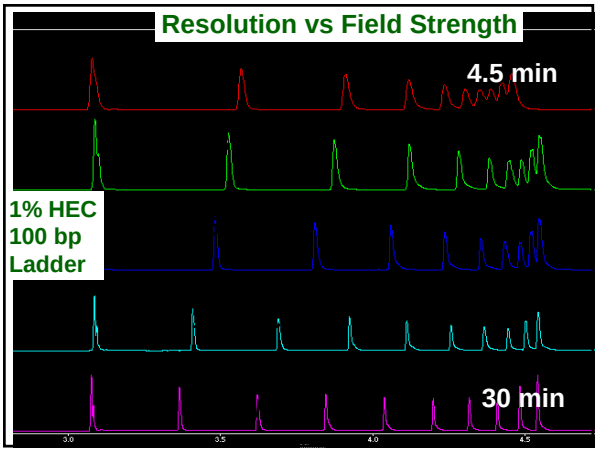
The electric field strength can influence the shape of the DNA molecule.



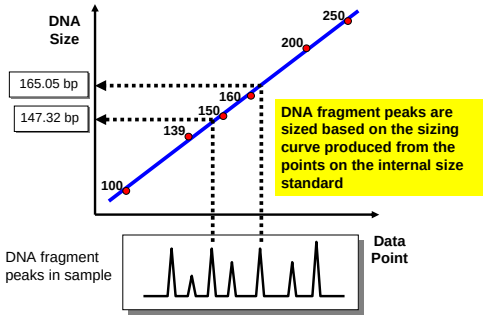
low *moderate* *high*

Optimal separations usually 180-200 V/cm



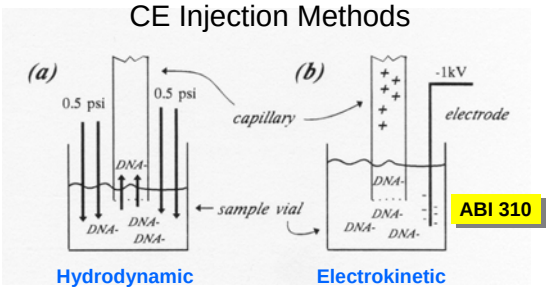


Process of Sizing DNA Fragments Using an Internal Standard



Injection

CE Injection Methods



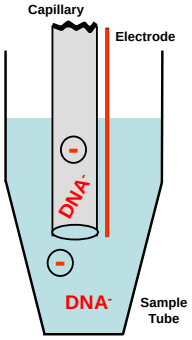
ABI 310

Hydrodynamic (pressure) **Electrokinetic (voltage)**

Ulfelder K. J.; McCord, B. R. (1996) Capillary Electrophoresis of DNA, In *Handbook of Capillary Electrophoresis* (Landers, J., ed.), CRC Press: NY, pp. 347-378.

Butler, J.M. (1997) Effects of sample matrix and injection on DNA separations. *Analysis of Nucleic Acids by Capillary Electrophoresis* (Heller, C., ed.), Vieweg: Germany, Chapter 5, pp. 125-134

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

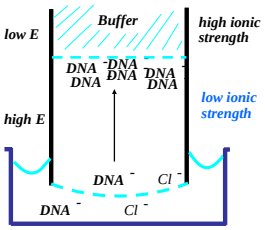
λ_{sample} is the sample conductivity

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

Typical Sample Preparation for ssDNA

1. Perform PCR with dye-labeled primers
2. Dilute 1 µL PCR product with 24 µL **deionized formamide**; add 1 µL ROX-labeled internal sizing standard
3. Denature 2 minutes at 95 °C with thermocycler
4. Cool to 4 °C in thermocycler or ice bath
5. Sample will remain denatured for at least 3 days

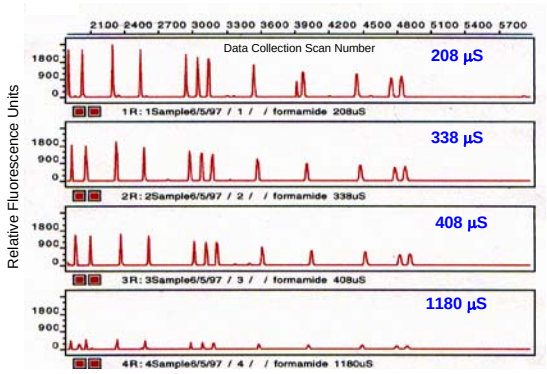


Figure 14.3, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

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*x*l 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

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)

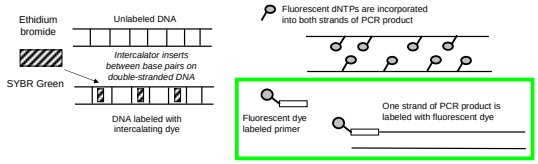
Filters determine which wavelengths of light are collected onto the CCD camera

Laser Used in ABI 310

- Argon Ion Laser
- 488 nm and 514.5 nm for excitation of dyes
- 10 mW power
- Lifetime ~5,000 hours (1 year of full-time use)
- Cost to replace ~$5,500
- Leads to highest degree of variability between instruments and is most replaced part
- Color separation matrix is specific to laser used on the instrument

Methods for Fluorescently Labeling DNA

- Intercalating Dyes (post-PCR)
- Dye-labeled nucleotide insertion during PCR
- Dye-labeled primer insertion during PCR

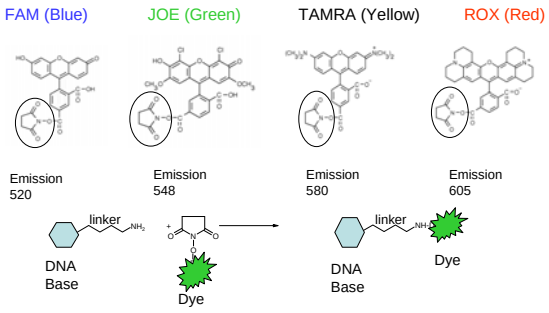


Butler, J.M. (2001) *Forensic DNA Typing*, Figure 10.2, ©Academic Press

Fluorescent Labeling of PCR Products

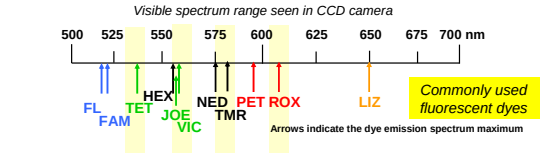
- Dyes are attached to one primer in a pair used to amplify a STR marker
- Dyes are coupled to oligonucleotides (primers) through NHS-esters and amine linkages on the 5'end of the primer: Dye-$(CH_2)_6$-primer
- Dye-labeled oligonucleotides are incorporated during multiplex PCR amplification giving a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

Amine Reactive Dyes used in Labeling DNA



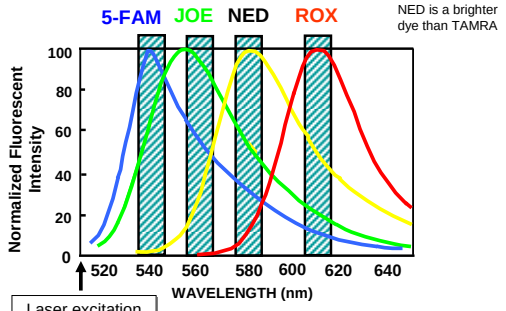
The succinimidyl ester reacts rapidly with amine linkers on DNA bases

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

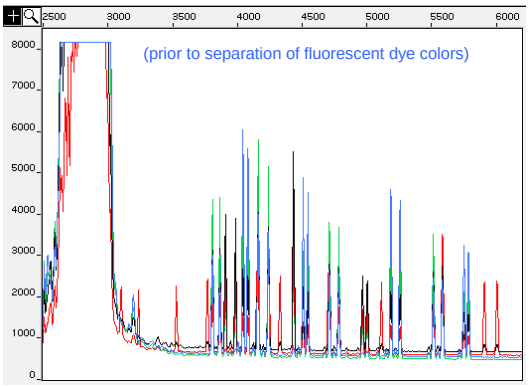


Butler, J.M. (2001) *Forensic DNA Typing*, Figure 10.4, ©Academic Press

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

Raw Data from the ABI Prism 310



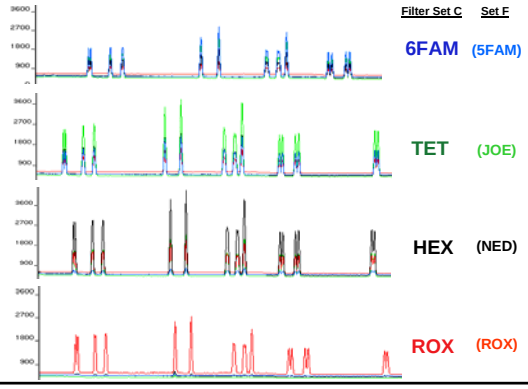
Why Make a Matrix?

The matrix is the solution to a problem:
What's the contribution at any given wavelength
(filter set) from each dye ?

There are 4 dyes

- **Remember algebra from high school?**
- **To solve a problem with 4 unknowns, you need 4 equations**

Matrix Standards (Raw Data)



For Example

$I_{540} = bx_b + gy_b + yz_b + rw_b$ **intensity of blue**
$I_{560} = bx_g + gy_g + yz_g + rw_g$ **intensity of green**
$I_{580} = bx_y + gy_y + yz_y + rw_y$ **intensity of yellow**
$I_{610} = bx_r + gy_r + yz_r + yw_r$ **intensity of red**

Where
b is the %blue labeled DNA
g is the %green labeled DNA, etc.

x,y,z,w are the numbers in the matrix (sensitivity to each color)

POP4STRMODF

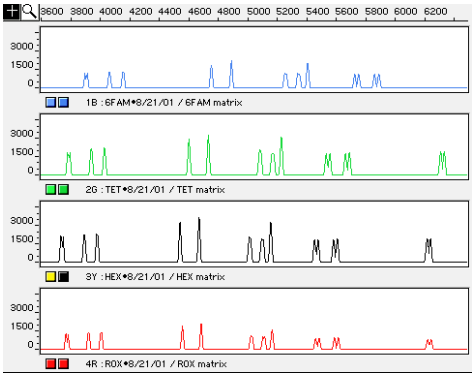
Reactions

	B	G	Y	R
B	1.0000	0.8502	0.1380	0.0009
G	0.8300	1.0000	0.7622	0.0051
Y	0.6416	0.8324	1.0000	0.1102
R	0.4493	0.6484	0.7851	1.0000

If you solve xyzw for each dye individually
Then you can determine dye contribution for any mixture

Matrix Standards

(After Color Separation)



Matrix File Table from an ABI 310

POP4STRMODF

Reactions

	B	G	Y	R
B	1.0000	0.8502	0.1380	0.0009
G	0.8300	1.0000	0.7622	0.0051
Y	0.6416	0.8324	1.0000	0.1102
R	0.4493	0.6484	0.7851	1.0000

The results of the calculation are in a matrix

(remember linear algebra?)

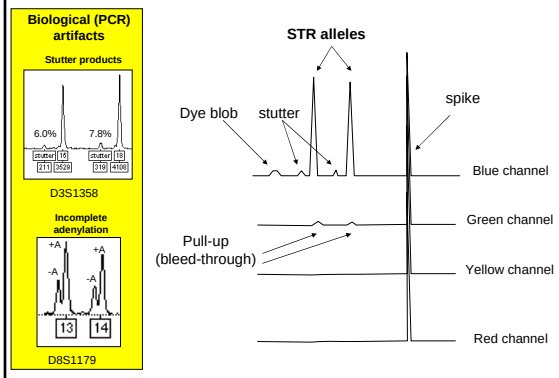
The values represent the percent spectral overlap from each dye

Values outside this range represent mixtures

Comments on Matrices/Spectral Calibration (Multi-Component Analysis)

- Make sure that the right filter set and matrix are applied when collecting data
- You can always apply another matrix to a sample collected on the ABI 310 but it must be run with the right filter set (matrix must be run first with ABI 3100)
- It is important to update matrices on a regular basis (depending on use) due to differences in laser power over time
- A good indication of when to run a new matrix is the observation of pull-up between dye colors when peaks are smaller than ~4,000 RFUs

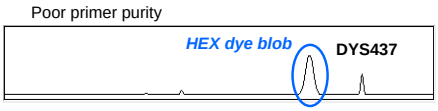
Deciphering Artifacts from the True Alleles



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

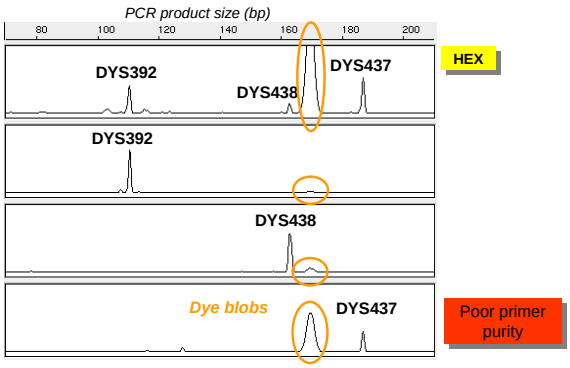
Dye Blobs ("Artifacts")

- 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)

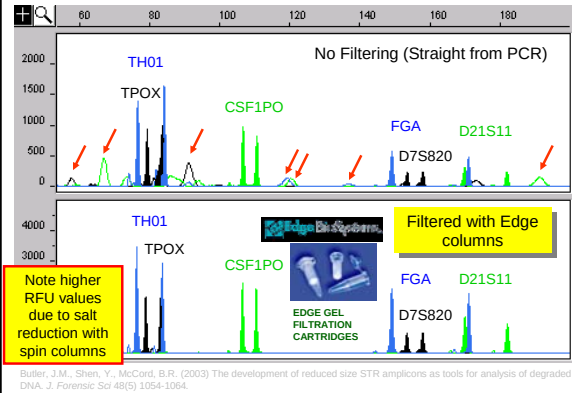


Dye Blob Problems with Some PCR Primers

Individual Y-STR Locus Amplifications

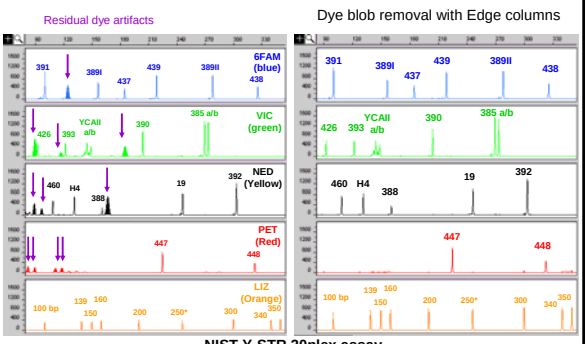


Removal of Dye Artifacts Following PCR Amplification



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.

Dye Artifacts Can Be Removed With Filtration



NIST Y-STR 20plex assay

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.*

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

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

- 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

Our Use of the ABI 3100

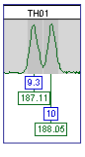
- Data collection software, version 1.0.1
- **POP-6** with 36 cm capillary array

- 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

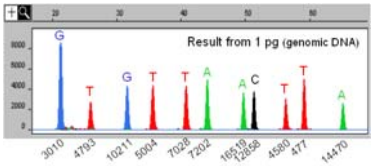
We can routinely get more than 400 runs per capillary array by not changing the polymer between applications

NIST ABI 3100 Analysis Using POP-6 Polymer

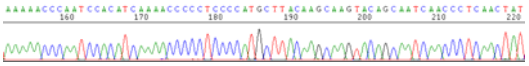
High Resolution STR Typing



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

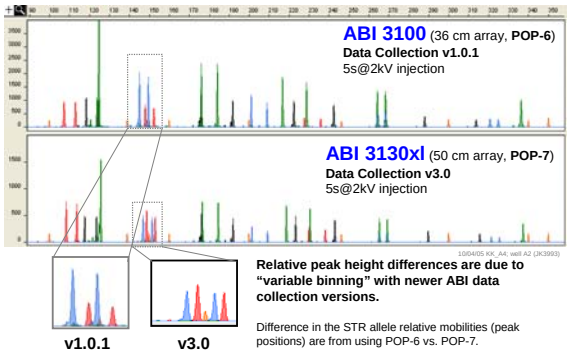


mtDNA Sequencing (HV1)



Comparison of ABI 3100 Data Collection Versions

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



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

Protocols Used for STR Typing

- Most forensic DNA laboratories follow PCR amplification and CE instrument protocols provided by the manufacturer
- Comments
  - Lower volume reactions may work fine and reduce costs
  - No heat denaturation/snap cooling is required prior to loading samples into ABI 310 or ABI 3100
  - Capillaries do not have to be thrown away after 100 runs
  - POP-4 polymer lasts much longer than 5 days on an ABI 310
  - **Validation does not have to be an overwhelming task**

Reduced Volume PCR Amplifications

Advantages

- **Lower cost** since kit contents are stretched
- Improved sensitivity perceived due to use of concentrated PCR products (since 1 uL out of a 5 uL reaction is 20% while 1 uL out of a 50 uL reaction is 2%)

Disadvantages

- Less volume of input DNA
  - **Tighter control (improved precision) required in DNA quantitation**
  - If low amount of DNA, then potential for allelic dropout (LCN conditions)
  - If PCR inhibitor is present, then less opportunity for dilution of inhibitor
- Evaporation impacts PCR amplification performance

Publications:

Gaines et al. *J Forensic Sci* 2002; 47(6):1224-1237. Reduced volume PCR amplification reactions using the AmpFlSTR Profiler Plus kit.
Leclair et al. *J Forensic Sci* 2003; 48(5):1001-1013. STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes.
Fregeau et al. *J Forensic Sci* 2003; 48(5):1014-1034. AmpFlSTR profiler Plus short tandem repeat DNA analysis of casework samples, mixture samples, and nonhuman DNA samples amplified under reduced PCR volume conditions (25 microL).

Identifiler 5 µL PCR Protocol

Identifiler PCR amplification was carried out on a GeneAmp® 9700 using 1 ng of DNA according to kit protocols with the exception of **reduced volume reactions** (5 µL instead of 25 µL) and **reduced cycles** (26 instead of 28).

Amplification products were diluted 1:15 in Hi-Di™ formamide and GS500-LIZ internal size standard (0.3 uL) and analyzed on the 16-capillary ABI Prism® 3100 Genetic Analyzer without prior denaturation of samples.

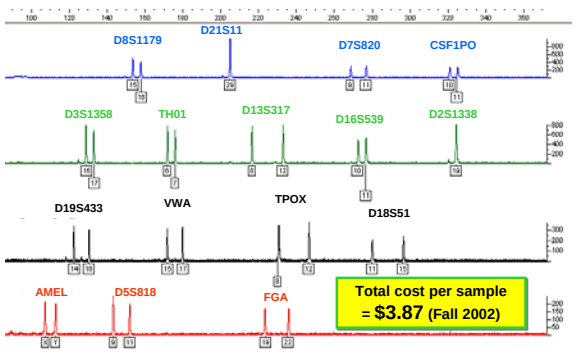
POP™-6 (3700 POP6) rather than POP™-4 was utilized for higher resolution separations.

Allele calls were made in Genotyper® 3.7 by comparison with kit allelic ladders using the Kazaam macro (20% filter).

Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *J Forensic Sci* 2003; 48(4):908-911.

Identifiler 5 µL PCR

(lower 3100 injection; 5s@2kV instead of 10s@3kV)



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…

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

Topics and Techniques for Forensic DNA Analysis
Continuing Education Seminar

Capillary Electrophoresis Troubleshooting

Utah DNA Training

Salt Lake City, UT
May 13-14, 2009

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

john.butler@nist.gov

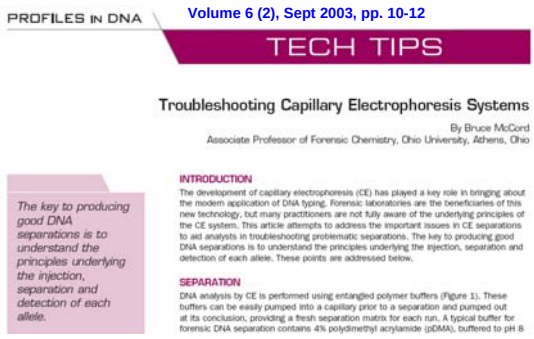
Troubleshooting Common Laboratory Problems

John M. Butler, PhD
National Institute of Standards and Technology

Bruce R. McCord, PhD
Florida International University

19th International Symposium on Human Identification
Hollywood, CA
October 16, 2008

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

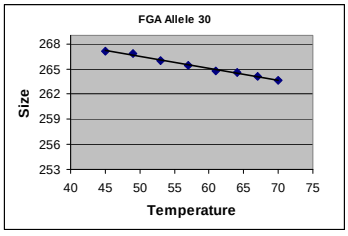
Outline for This Section

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

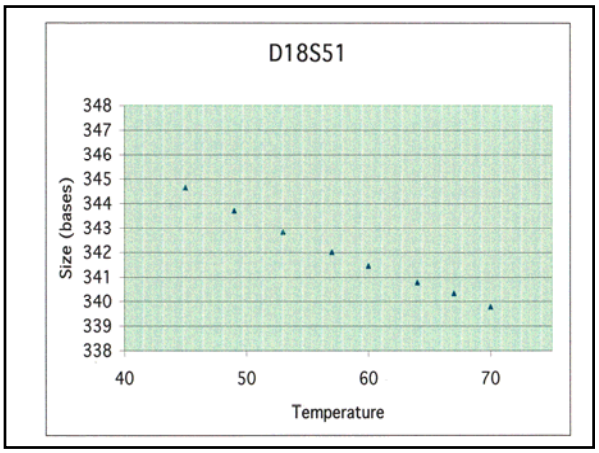
3. External Factors

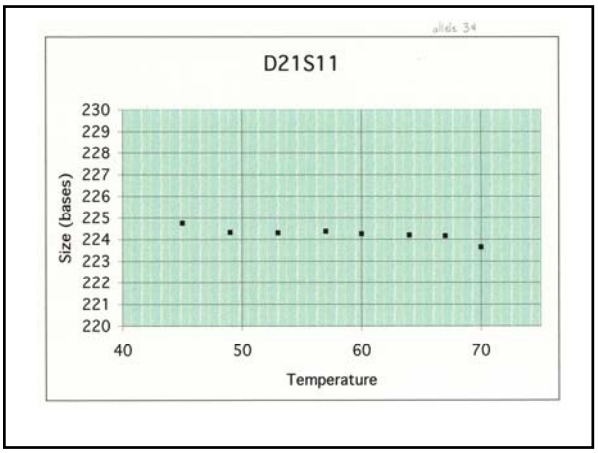
- Room temperature
  - Variations in room temperature can cause mobility shifts with band shifts and loss of calibration
  - Temperature is also important due to effects of high humidity on electrical conductance
- Cleanliness
  - Urea left in sample block can crystallize and catalyze further crystal 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

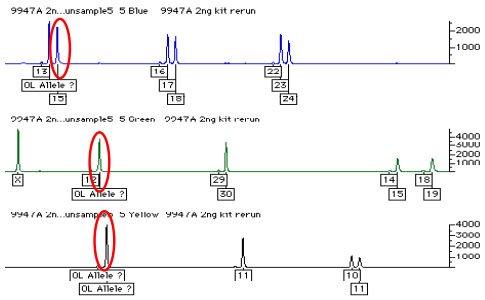


Slope is 0.14 bases/degree centigrade
Therefore a small change in temperature has a big effect
(A 1-2 degree shift in temperature of the heat plate can produce an OL allele)

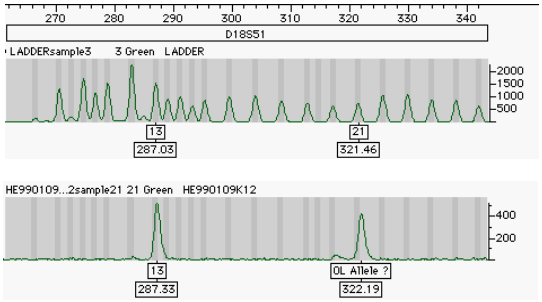




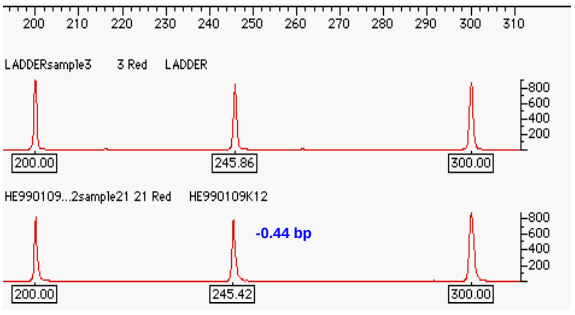
Temperature Effects: "OL" Alleles



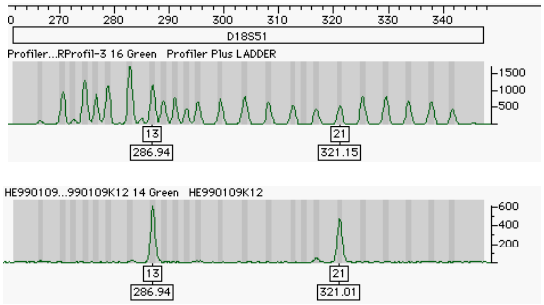
"OL Alleles"



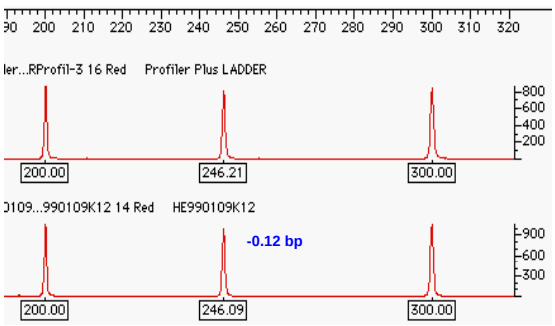
"OL alleles " - look at the 250 peak



"OL allele re-injected"



And the 250 peak...



Monitoring Room Temperature Over Time

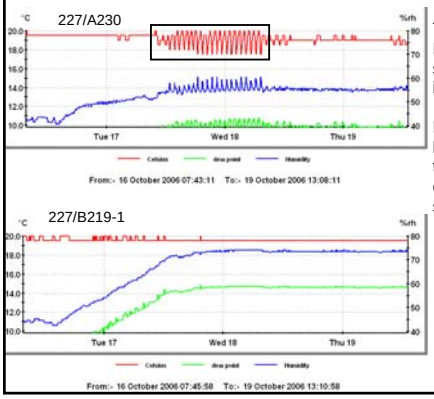
Temperature Probes

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

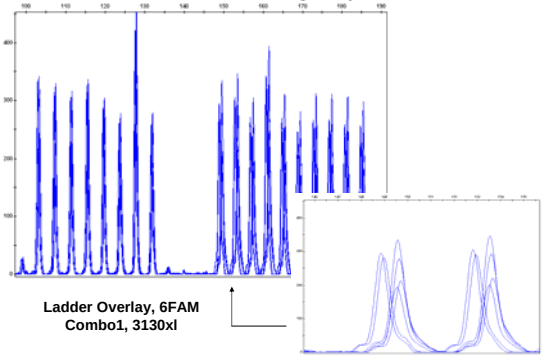
Monitoring Instrument Room Temperature Fluctuations



Temperature Monitoring of two separate instrument rooms.

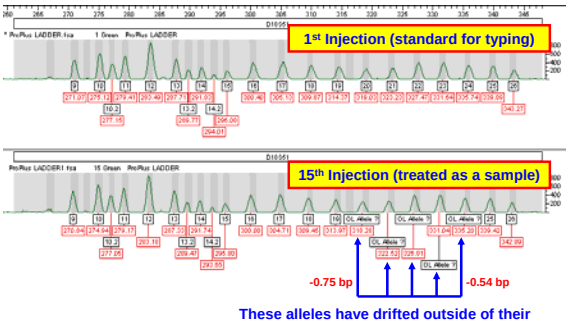
Box area is a 24 hour period where temperature control is not stable.

Poor Temperature Control Causes DNA Sizing Imprecision



Ladder Overlay, 6FAM Combo1, 3130xl

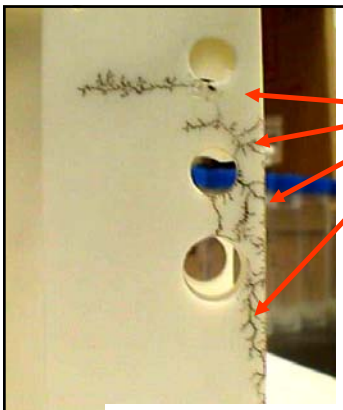
Use of Second Allelic Ladder to Monitor Potential Match Criteria Problems



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

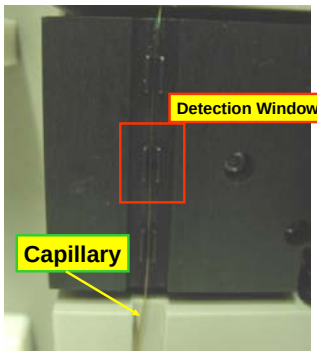


Keep Your System Clean!

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)

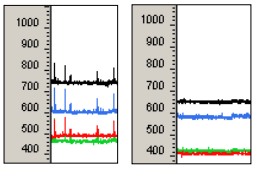
The Detection Window



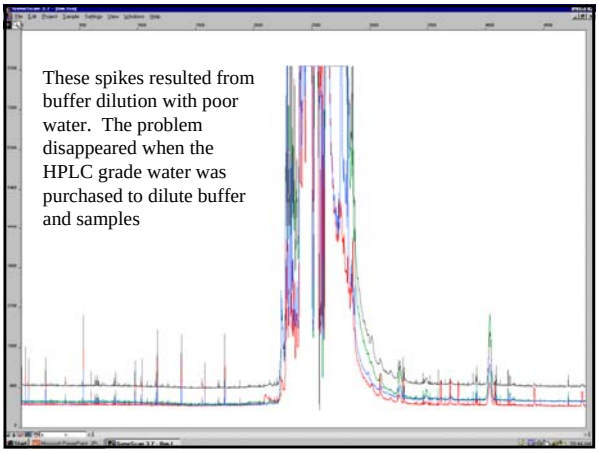
Make sure that the capillary window is lined up (if it is not, then no peaks will be seen)

Window may need to be cleaned with ethanol or methanol

Review Start of Raw Data Collection

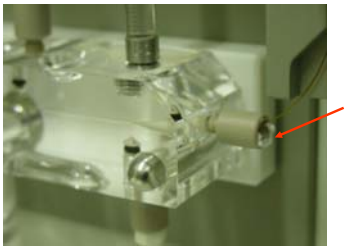


Little spikes indicate need to change buffer… check current



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



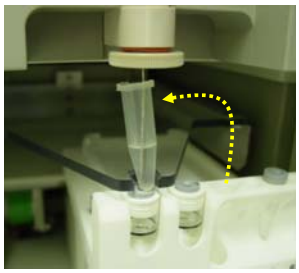
Urea crystals have formed due to a small leak where the capillary comes into the pump block

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



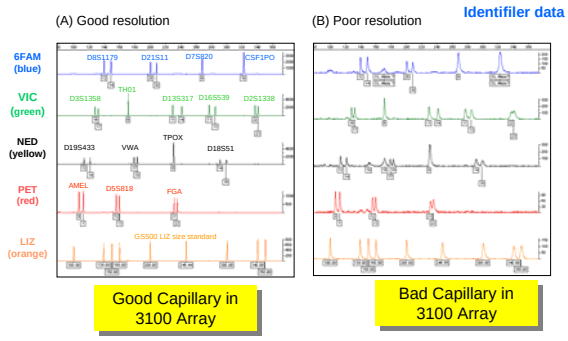
Remember that the water in the open tube will evaporate over time…

- 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

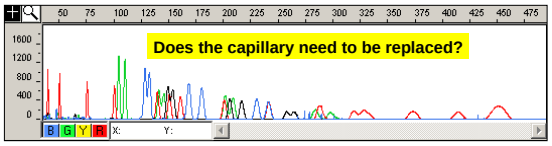
Capillary Meltdowns



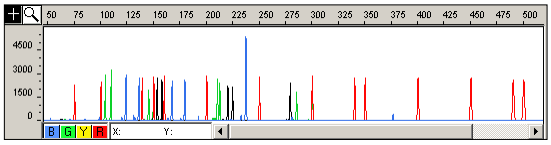
Butler, J.M., Buel, E., Crivellente, F., McCord, B.R. (2004) Forensic DNA typing by capillary electrophoresis: using the ABI Prism 310 and 3100 Genetic Analyzers for STR analysis. *Electrophoresis*, 25: 1397-1412.

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)

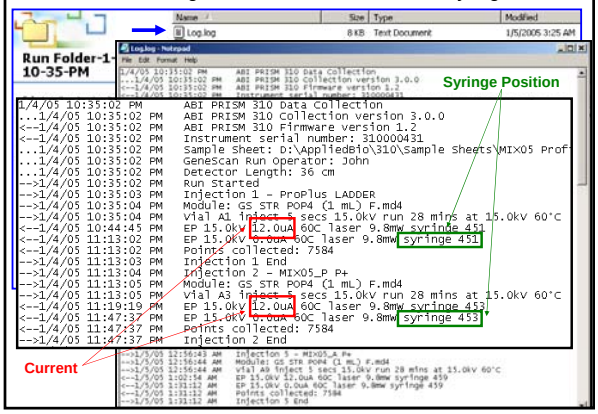
Measurement of Current

- 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)

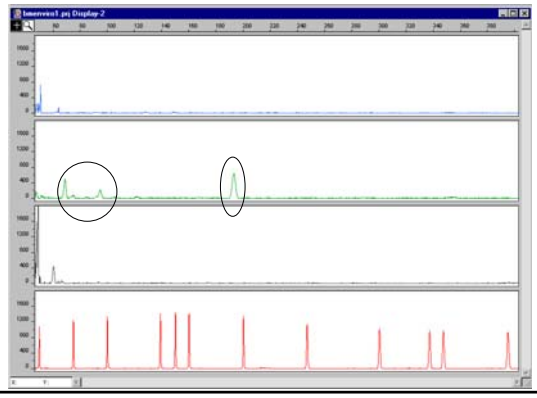
Syringe Travel

- 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.

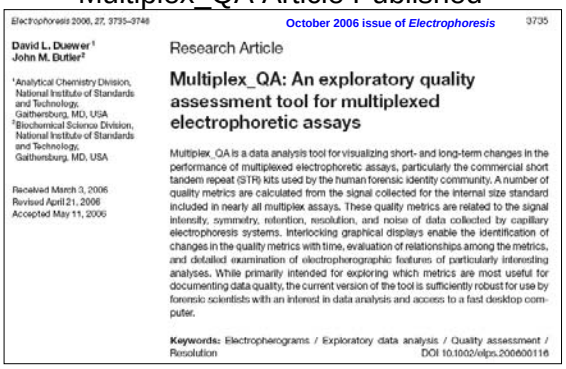
Conclusion:

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, 3735–3748 — 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

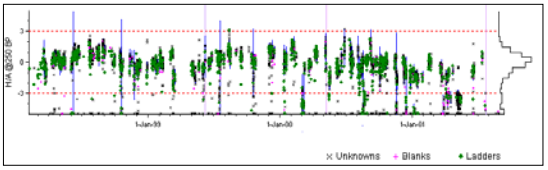
Multiplex_QA: An exploratory quality assessment tool for multiplexed electrophoretic assays

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 2008-DN-R-121 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

Topics and Techniques for Forensic DNA Analysis
Continuing Education Seminar

Y-STRs

Utah DNA Training

Salt Lake City, UT
May 13-14, 2009

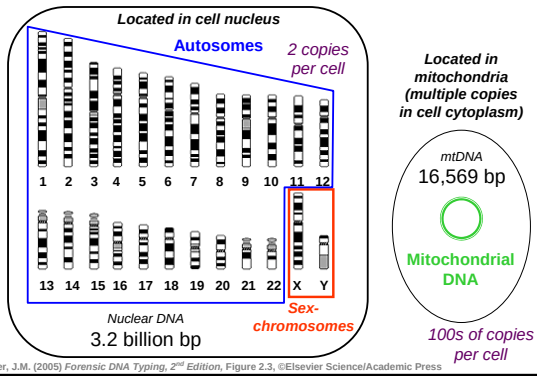
Dr. John M. Butler
National Institute of Standards and Technology
john.butler@nist.gov

Presentation Outline

- Why Y is of interest in human identity testing
- Y-STR markers and kits available
- Different population databases and statistics for reporting matches
- Mutation rates, duplications, and deletions and their impact on interpretation
- Value of additional Y-STR loci (beyond the Yfiler 17)

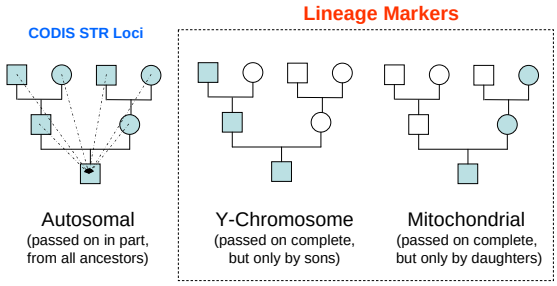
Human Genome

23 Pairs of Chromosomes + mtDNA



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 2.3, ©Elsevier Science/Academic Press

Different Inheritance Patterns



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.1, ©Elsevier Science/Academic Press

Role of Y-STRs and mtDNA

Compared to Autosomal STRs

- **Autosomal STRs provide a higher power of discrimination and <u>are the preferred method</u> whenever possible**
- **Due to capabilities for male-specific amplification**, Y-chromosome STRs (**Y-STRs**) can be useful in extreme female-male mixtures (e.g., when differential extraction is not possible such as fingernail scrapings)
- **Due to high copy number**, mitochondrial DNA (**mtDNA**) may be the only source of surviving DNA in highly degraded specimens or low quantity samples such as hair shafts

A mtDNA result is better than no result at all…

Lineage Markers: Y-STRs and mtDNA

<u>Advantages</u>

- Extend possible reference samples beyond a single generation (benefits missing persons cases and genetic genealogy)
- Family members have indistinguishable haplotypes unless mutations have occurred

<u>Disadvantages</u>

- Lower power of discrimination due to no genetic shuffling with recombination
- Family members have indistinguishable haplotypes unless mutations have occurred

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

Genetic Genealogy Companies

FamilyTreeDNA
http://www.familytreedna.com
http://www.dna-fingerprint.com

Sorenson Genomics
http://www.sorensongenomics.com

Relative Genetics
http://www.relativegenetics.com

oxford ancestors
http://www.oxfordancestors.com

DNA Heritage
http://www.dnaheritage.com

ETHNO ANCESTRY
http://www.ethnoancestry.com

GeoGene
http://www.geogene.com

The rapidly growing field of genetic genealogy is expanding the use of mtDNA and Y-STRs.

Perhaps the Real Reason Some Genetic Genealogy Is Performed…

"You don't look anything like the long haired, skinny kid I married 25 years ago. I need a DNA sample to make sure it's still you."

Summary of 2006 CODIS Survey Questions Regarding Y-STRs

171 labs

Questions #45a & #45b

- **Is your lab using or validating Y-STRs?**
  - **51 Yes** (30%)
    28 Yfiler, **15** PowerPlex Y, some both kits
  - **114** No
  - **6** no response

Summary of 2006 CODIS Survey Questions Regarding Y-STRs

171 labs

Question #50

- Y-STR data can be entered in CODIS similar to entering the current STR loci in CODIS. **Do you think CODIS should include Y-STR loci in Popstats calculations?**
  - Yes – 116 (68%)
  - No – 18
  - No response – 37

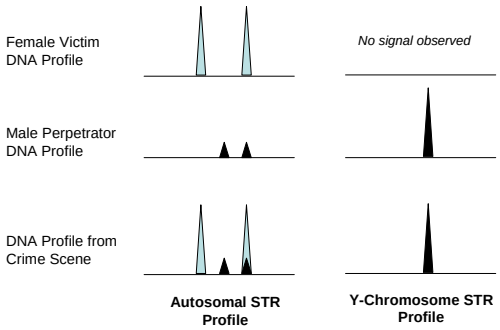
Past *Law & Order* episodes have discussed the "CODIS Y-STR database" and its capabilities for familial searching…

Value of Y-Chromosome Markers

J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition; Table 9.1

Application	Advantage
Forensic casework on sexual assault evidence	**Male-specific amplification** (can avoid differential extraction to separate sperm and epithelial cells)
Paternity testing	Male children can be tied to fathers in motherless paternity cases
Missing persons investigations	**Patrilineal male relatives may be used for reference samples**
Human migration and evolutionary studies	Lack of recombination enables comparison of male individuals separated by large periods of time
Historical and genealogical research	Surnames usually retained by males; can make links where paper trail is limited

Y-STRs can permit simplification of male DNA identification in sexual assault cases



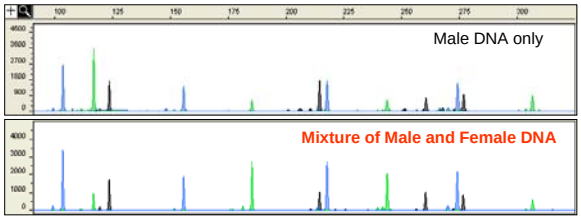
Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 9.2, ©Elsevier Science/Academic Press

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

Forensic Advantages of Y-STRs

- **Male-specific amplification** extends range of cases accessible to obtaining probative DNA results (e.g., fingernail scrapings, sexual assault without sperm)
- **Technical simplicity due to single allele profile**; can potentially recover results with lower levels of male perpetrator DNA because there is not a concern about heterozygote allele loss via stochastic PCR amplification; number of male contributors can be determined
- **Courts have already widely accepted STR typing**, instrumentation, and software for analysis (Y-STR markers just have different PCR primers)
- **Acceptance of statistical reports using the counting method** due to previous experience with mtDNA

Y-STRs Identify the Male Component even with Excess Female DNA



800X female DNA

Scenarios Where Y-STRs Can Aid Forensic Casework

- Sexual assaults by vasectomized or azoospermic males (no sperm left behind for differential extraction)
- Extending length of time after assault for recovery of perpetrator's DNA profile (greater than 48 hours)
- Fingernail scrapings from sexual assault victims
- Male-male mixtures
- Other bodily fluid mixtures (blood-blood, skin-saliva)
- Gang rape situation to include or exclude potential contributors
- Confirmation of amelogenin Y negative males

Confirmation of Amelogenin Negative Males


Normal AMEL X,Y male


AMEL X only

- **Often due to deletion of that entire region of the Y-chromosome rather than a primer binding site mutation**
- Most commonly seen in males of Indian subcontinent origin
- **Y-STRs help demonstrate that the AMEL X sample is really male**
- Chang *et al.* (2007) *Forensic Sci. Int.* 166: 115-120
  - 12/649 Malaysian males showed no AMEL Y
- Cadenas *et al.* (2007) *Forensic Sci. Int.* 166: 155-163
  - 5/77 Nepal males showed no AMEL Y

A new section on the NIST STRBase website will be created on this topic soon

Disadvantages of the Y-Chromosome

- Loci are not independent of one another and therefore rare random match probabilities cannot be generated with the product rule; must use haplotypes (combination of alleles observed at all tested loci)
- **Paternal lineages possess the same Y-STR haplotype** (barring mutation) and thus fathers, sons, brothers, uncles, and paternal cousins cannot be distinguished from one another
- **Not as informative as autosomal STR results**
  - **More like addition (10 + 10 + 10 = 30) than multiplication (10 x 10 x 10 = 1,000)**

What has happened in the past few years…

- "Full" Y-chromosome sequence became available in June 2003; over 350 Y-STR loci identified (only ~20 in 2000)
- **Selection of core Y-STR loci** (SWGDAM Jan 2003)
- **Commercial Y-STR kits released**
  - ~~Y PLEX 6,5,12 (2001-03),~~ **PowerPlex Y** (9/03), **Yfiler** (12/04)
- Many population studies performed and databases generated with thousands of Y-STR haplotypes
- Forensic casework demonstration of value of Y-STR testing along with court acceptance

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

History of Y-STR Marker Discovery

"Extended Haplotype"

1992 - DYS19 (Roewer et al.)
1994 - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)
1996 - DYS389I/II, DYS390, DYS391, DYS392, DYS393 (Roewer et al.)
1996 - DYF371, DYS425, DYS426 (Jobling et al.)
1997 - DYS288, DYS388 (Kayser et al.)
1998 - DYS385 a/b (Schneider et al.)

"Minimal Haplotype"

1999 - A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4 (White et al.)
2000 - DYS434, DYS435, DYS436, DYS437, DYS438, DYS439 (Ayub et al.)
2000 - G09411 (DYS462), G10123 (de Knijff unpublished)

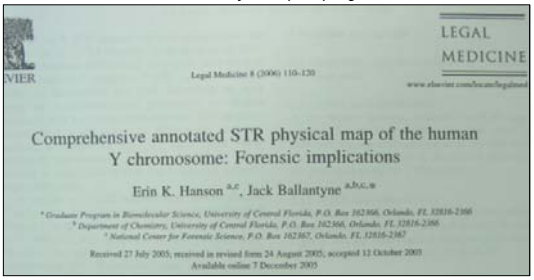
SWGDAM core

2001 - DYS441, DYS442 (Iida et al.)
2002 - DYS443, DYS444, DYS445 (Iida et al.); DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d (Redd et al.)
2002 – DYS468-DYS596 (129 new Y STRs; Manfred Kayser GDB entries)
2003 – DYS597-DYS645 (50 new Y STRs; Manfred Kayser GDB entries)
2004-2006 – DYS648-726 (GDB entries)

From J.M. Butler (2003) Recent developments in Y-STR and Y-SNP analysis. *Forensic Sci. Rev.* 15:91-111

Physical Map of the Human Y-Chromosome

Hanson, E.K. and Ballantyne, J. (2006) *Legal Med* 8: 110-120


LEGAL MEDICINE

Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications

Erin K. Hanson, Jack Ballantyne

Describe the precise location of 417 Y-STRs

They note that not all will be useful due to low genetic variation or high X-chromosome homology

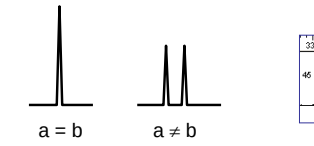
See also http://ncfs.ucf.edu/ystar/ystar.html

Y-STR Typing of Duplicated Regions

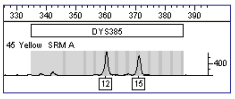
"multi-copy loci"

a repeat b repeat

Multiple primer binding sites occur giving rise to more than one PCR product for a given set of primers

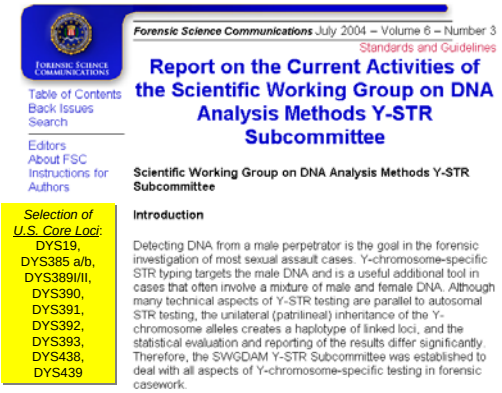

a = b a ≠ b

Y-PLEX™ 6 results



DYS385 a/b and YCAII a/b

Y-STR loci are often counted by the number of amplicons rather than the number of PCR primer pairs


Forensic Science Communications July 2004 – Volume 6 – Number 3
Standards and Guidelines

Report on the Current Activities of the Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee

Table of Contents
Back Issues
Search
Editors
About FSC
Instructions for Authors

Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee

Introduction

Detecting DNA from a male perpetrator is the goal in the forensic investigation of most sexual assault cases. Y-chromosome-specific STR typing targets the male DNA and is a useful additional tool in cases that often involve a mixture of male and female DNA. Although many technical aspects of Y-STR testing are parallel to autosomal STR testing, the unilateral (patrilineal) inheritance of the Y-chromosome alleles creates a haplotype of linked loci, and the statistical evaluation and reporting of the results differ significantly. Therefore, the SWGDAM Y-STR Subcommittee was established to deal with all aspects of Y-chromosome-specific testing in forensic casework.

Selection of U.S. Core Loci: DYS19, DYS385 a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

Core Y-STR Characteristics

11 PCR products
9 primer sets

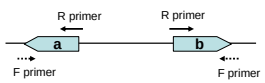
STR Marker	Position (Mb)	Repeat Motif	Allele Range	Mutation Rate
DYS393	3.17	AGAT	8-17	0.05%
DYS19	10.12	TAGA	10-19	0.20%
DYS391	12.54	TCTA	6-14	0.40%
DYS439	12.95	AGAT	8-15	0.38%
DYS389 I/II	13.05	[TCTG] [TCTA]	9-17 / 24-34	0.20%, 0.31%
DYS438	13.38	TTTTC	6-14	0.09%
DYS390	15.71	[TCTA] [TCTG]	17-28	0.32%
DYS385 a/b	19.19, 19.23	GAAA	7-28	0.23%
DYS392	20.97	TAT	6-20	0.05%

Positions in megabases (Mb) along the Y-chromosome were determined with NCBI build 35 (May 2004) using BLAT. Allele ranges represent the full range of alleles reported in the literature. Mutation rates summarized from YHRD (http://www.yhrd.org; accessed 6 Apr 2005).

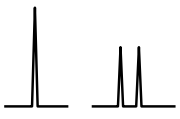
Butler, J.M. (2006) Genetics and genomics of core STR loci used in human identity testing. *J. Forensic Sci.*, 51(2): 253-265

(A) DYS385 a/b

Multi-Copy (Duplicated) Marker

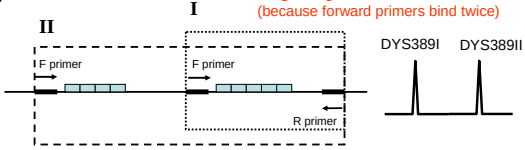


Duplicated regions are 40,775 bp apart and facing away from each other


a = b a ≠ b

(B) DYS389 I/II

Single Region but Two PCR Products
(because forward primers bind twice)



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 9.5, ©Elsevier Science/Academic Press

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

Primary Commercial Y-STR Kits

Allele size range and locus dye colors

PowerPlex® Y
Released by Promega Corporation in Sept 2003

100 bp 200 bp 300 bp 400 bp

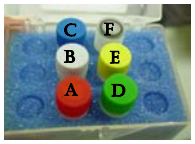
FL	DYS391	DYS389I	DYS439	DYS389II	
JOE	DYS438	DYS437	DYS19	DYS392	3 dye colors 12-plex PCR
TMR	DYS393	DYS390	DYS385a/b		

AmpFℓSTR® Yfiler™
Released by Applied Biosystems in Dec 2004

100 bp 200 bp 300 bp 400 bp

6-FAM	DYS456	DYS389I	DYS390	DYS389II		
VIC	DYS458	DYS19	DYS385a/b			4 dye colors 17-plex PCR
NED	DYS393	DYS391	DYS439	DYS635	DYS392	
PET	H4	DYS437	DYS438	DYS448		

Y-Chromosome Standard NIST SRM 2395



Human Y-Chromosome DNA Profiling Standard

- **5 male samples** + 1 female sample (neg. control)
- **100 ng of each (50 µL at ~2 ng/µL)**
- **22 Y STR markers sequenced**
- **9 additional Y STR markers typed**
- **42 Y SNPs typed** with Marligen kit

Certified for all loci in commercial Y-STR kits:

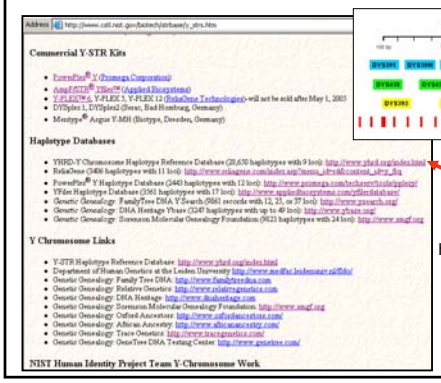
Y-PLEX 6
Y-PLEX 5
Y-PLEX 12
PowerPlex Y

SWGDAM recommended loci: DYS19, DYS385 a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

Y-filer - adds DYS635 (C4); now sequenced

Helps meet FBI Standard 9.5 (and ISO 17025)…traceability to a national standard

Y-Chromosome Information Resources on the NIST STRBase Website



Largest Y-STR Database
http://www.yhrd.org
41,965 haplotypes (9 loci)
14,835 haplotypes (11 loci)

Available Y-STR Loci, Kits and Databases

Loci	Grouping (# Loci)	Available Data
DYS19		http://www.YHRD.org **72,082 haplotypes** (>500 populations around the world)
DYS389I		
DYS389II		
DYS390	Minimal Haplotype (9)	
DYS391		
DYS392		NIJ-funded **US Database** at UCF: **17,216 haplotypes** **http://www.usystrdatabase.org/**
DYS393		
DYS385 a/b		
DYS438	SWGDAM Core (11)	http://www.YHRD.org 43,557 haplotypes
DYS439		
DYS437	PowerPlex Y (12)	Promega website: 4004 haplotypes
DYS448		
DYS456		Applied Biosystems website: **3561 haplotypes**
DYS458	Yfiler (17)	
DYS635		
GATA-H4		

~400 additional Y-STRs currently known
Hanson & Ballantyne, *Legal Med* 2006;8(2):110-20

Haplotype Databases for Y-STR Kits

http://www.promega.com/techserv/tools/pplexy/
http://www.appliedbiosystems.com/yfilerdatabase/

PowerPlex Y

1311 Caucasians
325 Asians
894 Hispanics
1108 African Americans
366 Native Americans

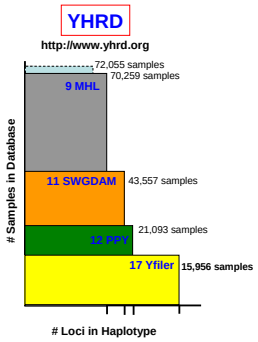
4,004 total
(as of March 2005)

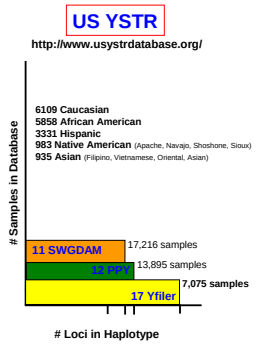
Yfiler

1276 Caucasians
330 Asians
597 Hispanics
985 African Americans
106 Native Americans
105 Filipino
59 Sub-Saharan Africans
103 Vietnamese

3,561 total
(as of December 2004)

Y-STR Database Content (as of May 2009)





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

SWGDAM Guidelines on Y-STR Interpretation

SWGDAM Y-STR Interpretation Guidelines

- **Approved July 15, 2008 by SWGDAM**
- Published in *Forensic Sci. Comm.* Jan 2009 issue


Forensic Science Communications

Y-chromosome Short Tandem Repeat (Y-STR) Interpretation Guidelines

Scientific Working Group on DNA Analysis Methods (SWGDAM)

Introduction

http://www.fbi.gov/hq/lab/fsc/backissu/jan2009/standards/2009_01_standards01.htm

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.1) Population Databases

- Loci on NRY should be considered linked as a single locus
- **Source of population database should be documented**
- Relevant population(s) for which the frequency will be estimated should be identified
- **Consolidated US Y-STR database should be used for population frequency estimation**

http://www.usystrdatabase.org

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.2) Haplotype Searches

- **Should be conducted using all loci for which results were obtained from the evidentiary sample**
- In cases where less information is obtained from the known sample, only those loci for which results were obtained from both the known and evidentiary sample should be used in the population database search

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.3) Haplotype Frequency Estimation

- **Counting method endorsed with application of a confidence interval** to correct for database size and sampling variation
- Reporting a haplotype count with a confidence interval is acceptable as a factual statement regarding observations in the database

	Haplotype observed previously	**Haplotype NOT observed previously**
Upper bound on 95% confidence interval	$p + 1.96\sqrt{\frac{(p)(1-p)}{n}}$	$1-(0.05)^{1/n}$

where p = x/n, n = database size, x = number of haplotypes in database

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.4) Y-STR Mixtures

- Calculations can be performed for probability of exclusion and likelihood ratios

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

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.5) Joint Match Probability

- **The product rule may be utilized** to combine the autosomal STR genotype match probability and Y-STR haplotype frequency information
- Citation to Walsh et al. (2008) Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Sci. Int.* 174: 234-238

SWGDAM Y-STR Interpretation Guidelines

Section 5. Statistical Interpretation

(5.6) Population Substructure

- Studies have shown that F_{st} values are very small for most populations
- **Use of the counting method that incorporates the upper bound estimate of the count proportion offers an appropriate and conservative statistical approach to evaluating the probative value of a match**

No need to use theta correction, but no discussion of partial profiles

Example Y-STR Haplotype

Core US Haplotype

- DYS19 – 14
- DYS389I – 13
- DYS389II – 29
- DYS390 – 24
- DYS391 – 11
- DYS392 – 14
- DYS393 – 13
- DYS385 a/b – 11,15
- DYS438 – 12
- DYS439 – 13

Matches by Databases

- YHRD (9 loci)
  - 7 matches in 27,773
- YHRD (11 loci)
  - 0 matches in 6,281
- ReliaGene (11 loci)
  - 0 matches in 3,403
- PowerPlex Y (12 loci)
  - 0 matches in 4,004
- Yfiler (17 loci)
  - 0 matches in 3,561

Y-Chromosome Haplotype Reference Database

www.YHRD.org

Release "15" from 2004-12-17 16:11:24

7 matches in 27,773
individuals from 236 worldwide populations

Minimal Haplotype Result

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 14
DYS393 – 13
DYS385 a/b – 11,15

Population	#	Metapopulation
Bogotá, Colombia [European]	1 / 147	Eurasian MP / European MP
Central Portugal	1 / 230	Eurasian MP / European MP
Cologne, Germany	1 / 135	Eurasian MP / European MP
Leipzig, Germany	1 / 661	Eurasian MP / European MP
Liguria, Italy	1 / 81	Eurasian MP / European MP
London, UK	1 / 285	Eurasian MP / European MP
Lyon, France	1 / 125	Eurasian MP / European MP

Frequency Estimate Calculations

In cases where a Y-STR profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

$p = X/N$

7 matches in 27,773

p = 7/27,773 = 0.000252 = **0.025%**

An upper bound confidence interval can be placed on the profile's frequency using:

$$p + 1.96\sqrt{\frac{(p)(1-p)}{N}}$$

$$0.000252 + 1.96\sqrt{\frac{(0.000252)(1-0.000252)}{27{,}773}}$$

= 0.000252 + 0.000187 = 0.000439
= **0.044% (~1 in 2270)**

When there is no match…

In cases where the profile has not been observed in a database, the upper bound on the confidence interval is

$1-\alpha^{1/N}$

0 matches in 4,004

where α is the confidence coefficient (0.05 for a 95% confidence interval) and N is the number of individuals in the database.

$1-\alpha^{1/N} = 1-(0.05)^{[1/4,004]} = 0.000748$
= **0.075% (~1 in 1340)**

If using database of 2,443, then the best you can do is 1 in 816

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

The Meaning of a Y-Chromosome Match

Conservative statement for a match report:

The Y-STR profile of the crime sample matches the Y-STR profile of the suspect (at *xxx* number of loci examined). Therefore, ***we cannot exclude the suspect*** as being the donor of the crime sample. In addition, we cannot exclude all patrilineal related male relatives and an unknown number of unrelated males as being the donor of the crime sample.

Y-STR Mutations

Mutations will impact kinship testing involving Y-STRs

(e.g., use of a paternal relative as a reference for a missing persons case)

NIST Work with Father-Son Samples

- Samples obtained from paternity testing laboratory as buccal swabs, extracted with DNA-IQ, quantified, diluted to 0.5 ng/uL
- To-date: **100 father-son pairs** of **U.S. Caucasian**, **African American**, **Hispanic**, and **Asian** (**800 samples**)
- **Verified** autosomal STR allele sharing **with Identifiler** (QC for gender and potential sample switches)
- Typed with Yfiler (17 Y-STRs) – **examined mutations**

Probability of Finding No Mutation or at Least One Mutation Between Two Y-STR Haplotypes in a Single Generation

Using average mutation rate of 0.28% (Kayser et al. AJHG 2000, 66:1580-1588)

# STRs	Prob. no mutation	Prob. at least one mutation	
1	0.99720000	0.00280000	
2	0.99440784	0.00559216	
3	0.99162350	0.00837650	
4	0.98884695	0.01115305	
5	0.98607818	0.01392182	
6	0.98331716	0.01668284	
7	0.98056387	0.01943613	
8	0.97781829	0.02218171	
9	0.97508040	0.02491960	
10	0.97235018	0.02764982	
11	0.96962760	0.03037240	
12	0.96691264	0.03308736	**3.3% with 12 Y-STRs**
…			
40	0.89390382	**0.10609618**	

Gusmão, L., Butler, J.M., et al. (2006) *Forensic Sci. Int.* 157:187-197

Separating Brothers with 47 Y-STRs

- Two suspected brothers (ZT79338 and ZT79339) are part of our ~660 U.S. sample dataset at NIST.
- Thus far, we have evaluated 47 Y-STR allele calls on these samples.
- **A mutation at DYS391 separates these individuals** (one contains allele 11 and the other allele 10).
- These samples share autosomal STR alleles and contain identical mtDNA sequences.

Y-STR Mutation Rates for the 17 Yfiler Loci

Yfiler kit loci	Literature Summary *			NIST Results			
Locus	Mutations	# Meioses	Mutation Rate	Mutations	# Meioses	Mutation Rate	TOTAL
DYS19	12	7272	0.165%	0	297	0.000%	**0.159%**
DYS389I	11	5476	0.201%	3	297	1.010%	**0.243%**
DYS389II	12	5463	0.220%	3	297	1.010%	**0.260%**
DYS390	16	6824	0.234%	1	293	0.341%	**0.239%**
DYS391	23	6702	0.343%	0	297	0.000%	**0.329%**
DYS392	4	6668	0.060%	0	297	0.000%	**0.057%**
DYS393	4	5456	0.073%	0	298	0.000%	**0.070%**
DYS385a/b	22	9980	0.220%	0	297	0.000%	**0.214%**
DYS438	1	2434	0.041%	0	297	0.000%	**0.037%**
DYS439	12	2409	0.498%	2	296	0.676%	**0.518%**
DYS437	5	2395	0.209%	0	296	0.000%	**0.186%**
DYS448	0	143	0.000%	0	294	0.000%	**<0.23%**
DYS456	1	143	0.699%	1	296	0.338%	**0.456%**
DYS458	3	143	2.098%	2	297	0.673%	**1.136%**
DYS635	3	1016	0.295%	3	298	1.007%	**0.457%**
GATA-H4	3	1179	0.254%	2	296	0.676%	**0.339%**

* Literature summary from www.YHRD.org and papers *in press*

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

Mutations Seen in 100 African American Father-Son Pairs

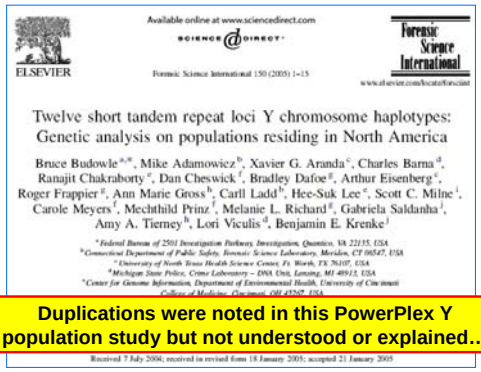
Ethnicity	Sample	locus	Allele (father)	Allele (child)	Comments
African American	65B	Y GATA H4	11	9	loss of 2 repeats
African American	46B	DYS389I and DYS389II	14,30	13,29	loss of 1 repeat
African American	58B	DYS389I and DYS389II	14,32	15,33	gain of 1 repeat
African American	18B	DYS390	24	23	loss of 1 repeat
African American	90B	DYS456	15	16	gain of 1 repeat
African American	16B	DYS458	18	19	gain of 1 repeat
African American	39B	DYS458	18	19	gain of 1 repeat
African American	16B	DYS635	23	22	loss of 1 repeat
African American	47B	DYS635	22	23	gain of 1 repeat
African American	72B	DYS635	22	23	gain of 1 repeat
African American	22B	DYS448	19,20	19,20	Duplication
African American	72B	DYS448	19,20	19,20	Duplication
African American	97B	DYS448	17.2,19,20	17.2,19,20	Triplication *
African American	33B	DYS389I and DYS389II			Deletion *
African American	33B	DYS439			Deletion *

Mutations in both DYS458 and DYS635 were observed in father and son 16B

Locus Duplication and Deletion

Events that impact Y-STR interpretation

PowerPlex Y Population Study


ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT

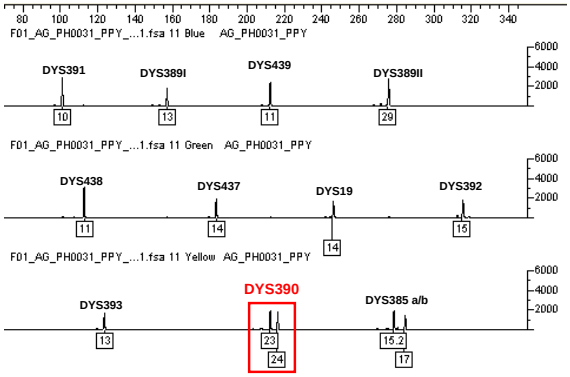
Forensic Science International

Twelve short tandem repeat loci Y chromosome haplotypes: Genetic analysis on populations residing in North America

Bruce Budowle, Mike Adamowicz, Xavier G. Aranda, Charles Barna, Ranajit Chakraborty, Dan Cheswick, Bradley Dafoe, Arthur Eisenberg, Roger Frappier, Ann Marie Gross, Carll Ladd, Hee-Suk Lee, Scott C. Milne, Carole Meyers, Mechthild Prinz, Melanie L. Richard, Gabriela Saldanha, Amy A. Tierney, Lori Viculis, Benjamin E. Krenke

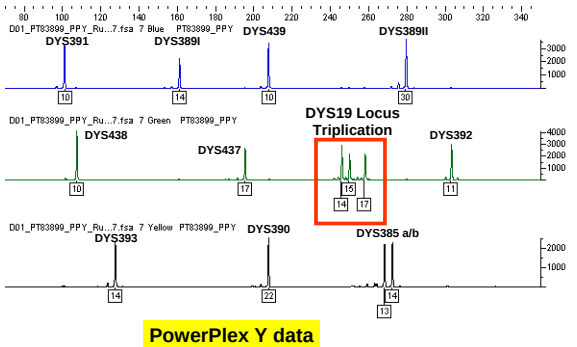
Duplications were noted in this PowerPlex Y population study but not understood or explained…

Duplication Observed at DYS390

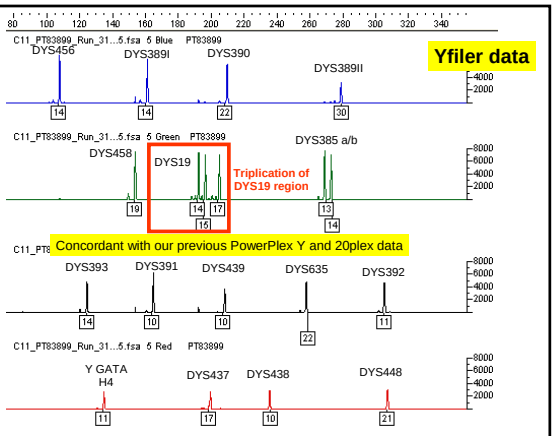


PowerPlex Y results on sample from Ann Marie Gross (MN BCA)

Locus Triplication at DYS19

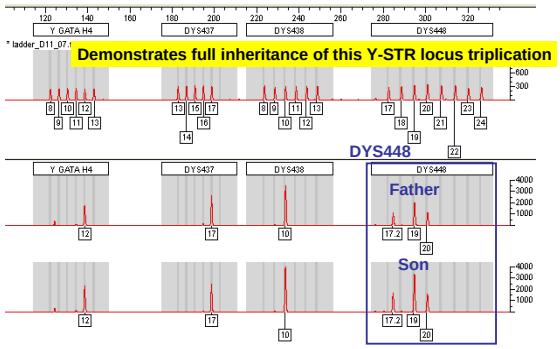


PowerPlex Y data



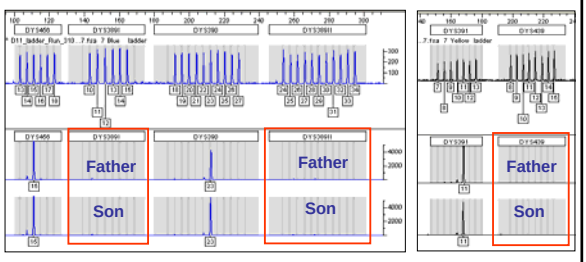
http://www.cstl.nist.gov/biotech/strbase/training.htm

DYS448 Triplication
Seen in Both Father and Son



DYS389I, DYS389II, DYS439 Deletions
Seen in Both Father and Son

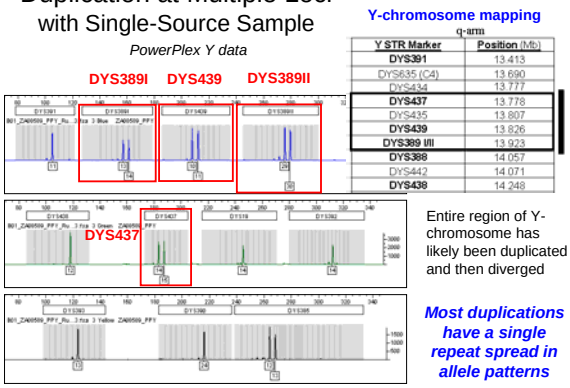
Yfiler data



Full inheritance of these Y-STR locus deletions

Duplication at Multiple Loci with Single-Source Sample

PowerPlex Y data



Y-chromosome mapping

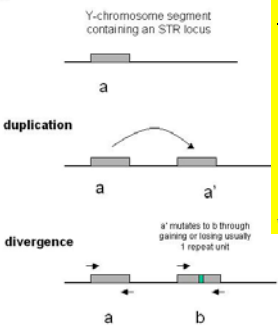
Y STR Marker	Position (Mb)
DYS391	13.413
DYS635 (C4)	13.690
DYS434	13.777
DYS437	13.778
DYS435	13.807
DYS439	13.826
DYS389 I/II	13.923
DYS388	14.057
DYS442	14.071
DYS438	14.248

Entire region of Y-chromosome has likely been duplicated and then diverged

Most duplications have a single repeat spread in allele patterns

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Duplication and Divergence Model



Locus	# dup*	>1 repeat
DYS19	23	2
DYS389I	5	0
DYS389II	9	2
DYS390	1	0
DYS391	3	1
DYS392	0	0
DYS393	3	0
DYS385a/b	17	0

*from www.yhrd.org, literature, and our work

92% have single repeat difference

Since single-step mutations are most common, then single repeat spacing in duplicated alleles is expected

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Deciphering between a Mixture of Multiple Males and Locus Duplication

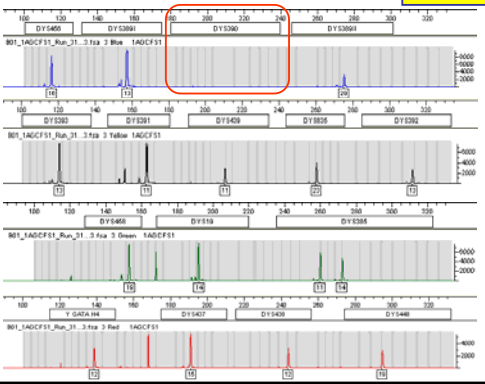
- ***Note the number of loci containing >1 allele (other than multi-copy DYS385)***
- ***Consider relative position on the Y-chromosome if multiple loci have two alleles***
- ***See if repeat spread is >1 repeat unit***
- ***Examine DYS385 for presence of >2 alleles***

Locus duplication along the Y-chromosome is in many ways analogous to heteroplasmy in mitochondrial DNA, which depending on the circumstances can provide greater strength to a match between two DNA samples.

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

CFS-1 with Y filer

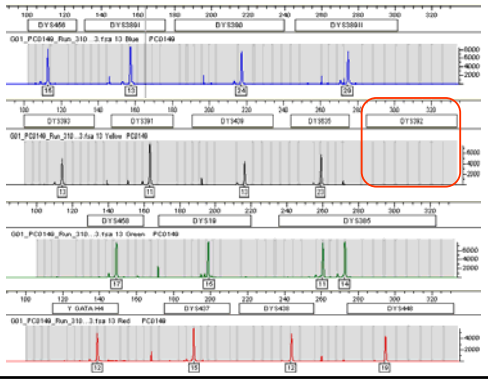
DYS390 is deleted



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

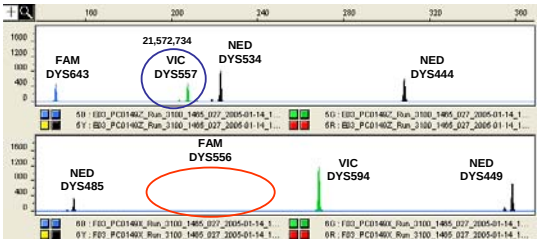
Sample PC0149 with Yfiler

DYS392 is deleted



PC0149 with Additional Y-STRs

One of the closest available loci fails



DYS556 ~32,000 bp away from DYS392 is **missing**

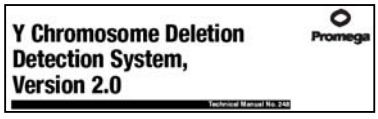
DYS557 ~600,000 bp away from DYS392 is present

Deletions of some Y-STRs can be an inadvertent diagnosis of male infertility

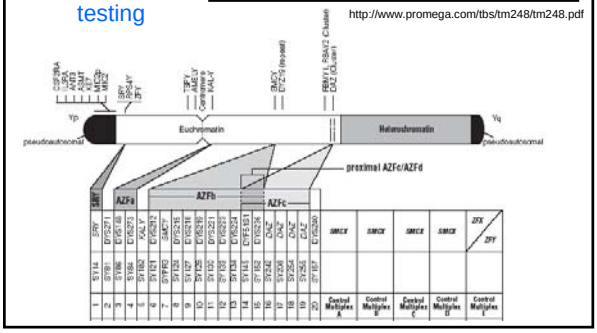
King *et al.* (2005) Inadvertent diagnosis of male infertility through genealogical DNA testing. *J. Med. Genet*. 42:366-368

- ***AZFa* deletion** (<1 in 100,000 men): expected to lack **DYS389I/II, DYS437, DYS438, DYS439**
- ***AZFb* deletion** (very rare): expected to lack **DYS385 and DYS392**
- ***AZFc* deletion** (1 in 4,000 men): expected to lack DYS464
- Possible that "incomplete" haplotypes are not being submitted to the Y-STR haplotype databases
- **Thus, Y-STRs are not neutral with respect to fertility information**

Promega sells a Y-deletion test for infertility testing



http://www.promega.com/tbs/tm248/tm248.pdf



Practical Information on Y Deletions

- If DYS458 is deleted in Yfiler, then your sample is likely to lack an Amelogenin Y amplicon as DYS458 and AMEL Y are 1.13 Mb apart on the short arm of the human Y-chromosome
  - Chang *et al.* (2007) *Forensic Sci. Int*. 166: 115-120
- Many Y-chromosomes are more complicated than originally thought!

Value of Additional Loci

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

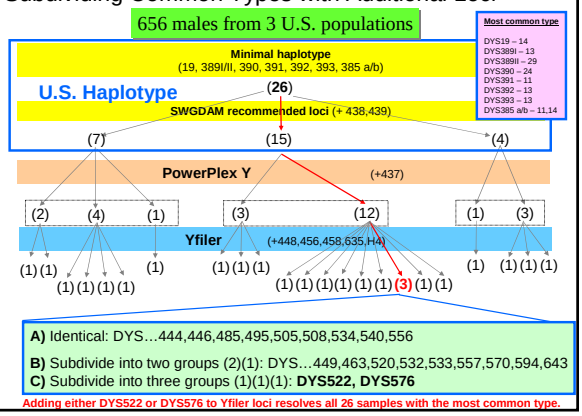
Going Beyond Commercial Y-STR Kits

- Most forensic DNA laboratories (certainly in the U.S.) will **only use commercially available kits** due to quality control issues
- Using these kits as a starting point, **are there additional loci that would be beneficial in separating samples with common types**, which could be advocated to companies for possible future adoption in Y-STR kits?
- Is it possible to regularly **resolve individuals from the same paternal lineage** (e.g., fathers and sons) if enough Y-STRs are examined?

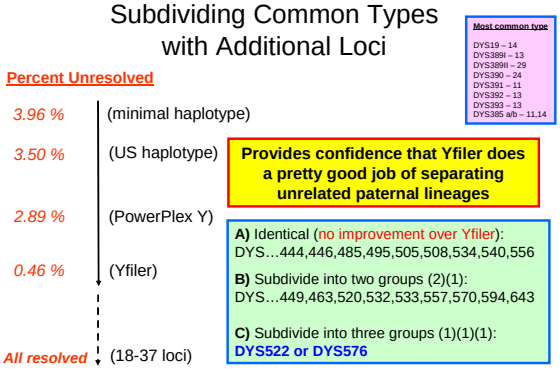
Data Set Used to Examine Common Types

- Yfiler kit (**17 Y-STR loci**) run on all NIST male U.S. population samples
  - makes up ~20% of Applied Biosystems database
  - **submitted to the YHRD**
- Additional **20 Y-STR loci** run on full set of NIST population samples *(and several less polymorphic ones only on subset of samples)*
  - Butler, J.M., Decker, A.E., Vallone, P.M., Kline, M.C. (2006) Allele frequencies for 27 Y-STR Loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci. Int.* 156:250-260.

Subdividing Common Types with Additional Loci



Subdividing Common Types with Additional Loci



# times haplotype observed	9 MHL
1	429
2	34
3	13
4	4
5	3
6	1
7	1
8	1
9	2
10	.
11	1
12	.
13	1
15	.
26	1
HD	0.996644
%DC	0.748476
# HT	491

429 of the 656 had a unique haplotype with the MHL loci, 34 sample haplotypes were observed twice in the sample set, 13 sample haplotypes were observed three times, etc.

With the 9 loci of the minimal haplotype (MHL) run on 656 samples, 26 samples had the most common type

Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM
1	429	486
2	34	33
3	13	10
4	4	6
5	3	1
6	1	1
7	1	2
8	1	.
9	2	.
10	.	1
11	1	.
12	.	.
13	1	.
15	.	1
26	1	.
HD	0.996644	0.998529
%DC	0.748476	0.824695
# HT	491	541

With the 11 loci of the SWGDAM haplotype run on 656 samples, 15 samples had the most common type

Total = 656 samples

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

# times haplotype observed	9 MHL	11 SWGDAM	12 PPY
1	429	486	505
2	34	33	34
3	13	10	14
4	4	6	3
5	3	1	2
6	1	1	.
7	1	2	1
8	1	.	.
9	2	.	.
10	.	1	.
11	1	.	.
12	.	.	**1**
13	1	.	.
15	.	**1**	.
26	**1**	.	.
HD	0.996644	0.998529	0.999064
%DC	0.748476	0.824695	0.853659
# HT	491	541	560

With the 12 loci of the PowerPlex Y haplotype (PPY) run on 656 samples, 12 samples had the most common type

Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM	12 PPY	17 Yfiler
1	429	486	505	626
2	34	33	34	12
3	13	10	14	2
4	4	6	3	.
5	3	1	2	.
6	1	1	.	.
7	1	2	1	.
8	1	.	.	.
9	2	.	.	.
10	.	1	.	.
11	1	.	.	.
12	.	.	**1**	.
13	1	.	.	.
15	.	**1**	.	.
26	**1**	.	.	.
HD	0.996644	0.998529	0.999064	0.999916
%DC	0.748476	0.824695	0.853659	0.97561
# HT	491	541	560	640

With the 17 loci in Yfiler across the 656 samples, there are 626 unique haplotypes, 12 haplotypes that were observed twice and 2 haplotypes that were observed three times

Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM	12 PPY	17 Yfiler	ALL 37
1	429	486	505	626	652
2	34	33	34	12	2
3	13	10	14	2	.
4	4	6	3	.	.
5	3	1	2	.	.
6	1	1	.	.	.
7	1	2	1	.	.
8	1	.	.	.	.
9	2	.	.	.	.
10	.	1	.	.	.
11	1	.	.	.	.
12	.	.	**1**	.	.
13	1	.	.	.	.
15	.	**1**	.	.	.
26	**1**	.	.	.	.
HD	0.996644	0.998529	0.999064	0.999916	0.999991
%DC	0.748476	0.824695	0.853659	0.97561	0.996951
# HT	491	541	560	640	654

When all 37 loci (Yfiler + 20 new loci) are run on 656 samples, only two haplotypes are observed twice

These two sets of three unseparated Yfiler types will be examined next

Total = 656 samples

Subdividing Unresolved Yfiler Haplotypes (1)

Most Common Type

Sample Info	DYS 19	DYS 385a/b	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	H4
MT97185	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12
ZT79333	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12
TT51702	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12

Locus	MT97185	ZT79333	TT51702
~~DYS444~~	~~12~~	~~12~~	~~12~~
~~DYS446~~	~~13~~	~~13~~	~~13~~
DYS449	30	30	31
DYS463	24	24	23
~~DYS485~~	~~15~~	~~15~~	~~15~~
~~DYS495~~	~~16~~	~~16~~	~~16~~
~~DYS505~~	~~12~~	~~12~~	~~12~~
~~DYS508~~	~~11~~	~~11~~	~~11~~
DYS520	21	22	21
DYS522	10	12	11

Locus	MT97185	ZT79333	TT51702
DYS532	14	14	13
DYS533	13	12	13
~~DYS534~~	~~15~~	~~15~~	~~15~~
~~DYS540~~	~~12~~	~~12~~	~~12~~
~~DYS556~~	~~11~~	~~11~~	~~11~~
DYS557	15	17	17
DYS570	16	17	17
DYS576	17	20	18
DYS594	9	10	10
DYS643	10	11	10

Either DYS522 or DYS576 will fully resolve all three of these samples

Subdividing Unresolved Yfiler Haplotypes(2)

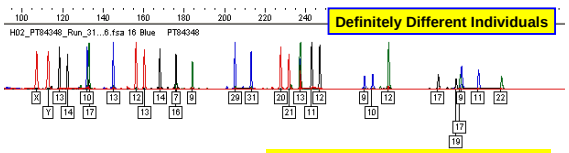
Sample Info	DYS 19	DYS 385a/b	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	H4
PT83904	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12
PT84348	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12
ZT80369	13	13,14	15	31	24	9	11	13	10	10	14	20	16	18	21	12

Locus	PT83904	PT84348	ZT80369
~~DYS444~~	~~12~~	~~12~~	~~12~~
~~DYS446~~	~~12~~	~~12~~	~~12~~
~~DYS449~~	~~31~~	~~31~~	~~31~~
~~DYS463~~	~~18~~	~~18~~	~~18~~
~~DYS485~~	~~15~~	~~15~~	~~15~~
~~DYS495~~	~~12~~	~~12~~	~~12~~
~~DYS505~~	~~11~~	~~11~~	~~11~~
~~DYS508~~	~~11~~	~~11~~	~~11~~
~~DYS520~~	~~19~~	~~19~~	~~19~~
~~DYS522~~	~~12~~	~~12~~	~~12~~

Locus	PT83904	PT84348	ZT80369
~~DYS532~~	~~14~~	~~14~~	~~14~~
~~DYS533~~	~~11~~	~~11~~	~~11~~
DYS534	16	17	17
~~DYS540~~	~~11~~	~~11~~	~~11~~
~~DYS556~~	~~12~~	~~12~~	~~12~~
~~DYS557~~	~~18~~	~~18~~	~~18~~
~~DYS570~~	~~22~~	~~22~~	~~22~~
~~DYS576~~	~~18~~	~~18~~	~~18~~
~~DYS594~~	~~11~~	~~11~~	~~11~~
~~DYS643~~	~~12~~	~~12~~	~~12~~

PT84348 and ZT80369 are unresolved even with 37 Y-STRs

Identifiler Results from Two Unresolved Hispanic Males Following Typing with 37 Y-STR Loci



Definitely Different Individuals

Not possible to tell if these samples are from brothers or paternal cousins

No allele sharing at 6 out of 15 STRs examined

Sample	D8	D21	D7	CSF	D3	TH	D13	D16	D2	D19	VWA	TPO	D18	Am	D5	FGA
PT84348	10,13	29,31	9,10	9,11	17,17	7,9	13,13	12,12	17,22	13,14	14,16	11,12	17,19	X,Y	12,13	20,21
ZT80369	14,15	28,31	10,11	11,12	16,18	7,9	8,11	12,13	23,25	14,15	14,15	8,11	13,15	X,Y	12,12	22,25

PT84348 and ZT80369 are unresolved even with 37 Y-STRs

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

Summary on Subdividing Common Types

- 640 haplotypes were observed in the 656 U.S. population samples with the Yfiler loci: 626 were unique, 2 were observed 3 times, and 12 haplotypes were observed twice.
- With the addition of 20 new Y-STR loci, all but two sample pairs are resolved.
- In this sample set, the 7 Y-STRs (DYS532, **DYS522, DYS576**, DYS570, DYS505, DYS449, DYS534) have the same ability to resolve the sample haplotypes as all 20 new loci.
- **These 7 loci will be the focus of future studies and multiplex assays**.

NIST Activities with Y-STRs

- SRM 2395 (Human Y Chromosome Standard)
  - http://www.cstl.nist.gov/biotech/strbase/SRM2395.htm
- Characterized duplications and deletions
  - Butler et al. (2005) *J. Forensic Sci.* 50(4): 853-859
- Sequenced variant alleles
  - http://www.cstl.nist.gov/biotech/strbase/STRseq.htm
- Supplied ~20% of Yfiler 3561 database
  - http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm
- Measured mutation rates with Yfiler loci
  - Decker et al. (2008) *FSI Genetics* 2(3): e31-e35

26 publications since 2001 on NIST Y-chromosome work

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

Acknowledgments

Funding from interagency agreement 2008-DN-R-121 between the National Institute of Justice 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

Tom Reid (DNA Diagnostics Center) – supplying the father-son samples for mutation rate analysis

Capillary Electrophoresis Literature Listing *Prepared by John Butler (NIST) – May 2009*

General Information

Butler, J.M. et al. (2004) Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25: 1397-1412.

McCord, B.R. (2003) Troubleshooting capillary electrophoresis systems. *Profiles in DNA* 6(2): 10-12; Available at: http://www.promega.com/profiles/.

Early Work

Butler, J.M. et al. (1994) Rapid analysis of the short tandem repeat HUMTH01 by capillary electrophoresis. *BioTechniques* 17: 1062-1070.

Butler, J.M. et al. (1995) Application of duel internal standards for precise sizing of polymerase chain reaction products using capillary electrophoresis. *Electrophoresis* 16: 974-980.

Wang, Y. et al. (1995) Rapid sizing of short tandem repeat alleles using capillary array electrophoresis and energy-transfer fluorescent primers. *Anal. Chem.* 67: 1197-1203.

Isenberg, A.R. et al. (1998) Analysis of two multiplexed short tandem repeat systems using capillary electrophoresis with multiwavelength florescence detection. *Electrophoresis* 19: 94-100.

Mansfield, E.S. et al. (1996) Sensitivity, reproducibility, and accuracy in short tandem repeat genotyping using capillary array electrophoresis. *Genome Res.* 6: 893-903.

ABI Prism 310 Genetic Analyzer

Rosenblum, B.B. et al. (1997) Improved single-strand DNA sizing accuracy in capillary electrophoresis. *Nucleic Acids Res.* 25: 3925-3929.

Wenz, H.M. et al. (1998) High-precision genotyping by denaturing capillary electrophoresis. *Genome Res.* 8: 69-80.

Buel, E. et al. (1998) Capillary electrophoresis STR analysis: Comparison to gel-based systems. *J. Forensic Sci.* 43: 164-170.

Lazaruk, K. et al. (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19: 86-93.

Moretti, T.R. et al. (2001) Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J. Forensic Sci.* 46: 647-660.

Moretti, T.R. et al. (2001) Validation of STR typing by capillary electrophoresis. *J. Forensic Sci.* 46: 661-676.

ABI 3100 and Other Capillary Array Systems

Gill, P. et al. (2001) Sizing short tandem repeat alleles in capillary array gel electrophoresis instruments. *Electrophoresis* 22: 2670-2678.

Sgueglia, J.B. et al. (2003) Precision studies using the ABI prism 3100 genetic analyzer for forensic DNA analysis. *Anal. Bioanal. Chem.* 376: 1247-1254.

Stewart, J.E. et al. (2003) Evaluation of a multicapillary electrophoresis instrument for mitochondrial DNA typing. *J. Forensic Sci.* 48: 571-580.

Koumi, P. et al. (2004) Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. *Electrophoresis* 25: 2227-2241.

Energy Transfer Dyes

Yeung, S.H.I. et al. (2008) Fluorescence energy transfer-labeled primers for high-performance forensic DNA profiling. *Electrophoresis* 29: 2251-2259.

DNA Separation Mechanisms

Barron, A.E. and Blanch, H.W. (1995) DNA separations by slab gel and capillary electrophoresis. *Separation Purification Meth.* 24: 1-118.

Useful References on Y-Chromosome Markers and Assays

Prepared by John Butler (May 2009)

Y-STR Haplotype Databases (see http://www.cstl.nist.gov/biotech/strbase/y_strs.htm)

Y-Chromosome Haplotype Reference Database (YHRD):http://www.yhrd.org (Feb 2009: 72,082 total haplotypes – 15,956 Yfiler haplotypes)

U.S. Y-STR Database: http://www.usystrdatabase.org/ (Mar 2009: 17,216 total haplotypes – 7,075 Yfiler haplotypes)

Genealogy Y-STR Database: http://www.ysearch.org/

Ballantyne, J. et al. (2006) Creating and managing effective Y-STR databases. *Profiles in DNA* 9(2): 10-13.

Willuweit, S. and Roewer, L. (2007) Y chromosome haplotype reference database (YHRD): Update. *Forensic Sci. Int. Genet.* 1: 83-87.

Fatolitis, L. and Ballantyne, J. (2008) The US Y-STR database. *Profiles in DNA* 11(1): 13-14.

ISFG Recommendations (and NIST recommendations on nomenclature)

Gill, P. et al. (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Forensic Sci. Int.* 124: 5-10.

Gusmão, L. et al. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci. Int.* 157:187-197.

Butler, J.M. et al. (2008) Addressing Y-chromosome short tandem repeat (Y-STR) allele nomenclature. *Journal of Genetic Genealogy* 4(2): 125-148.

STR Assays and Kits

Butler, J.M., Schoske, R., Vallone, P.M., Kline, M.C., Redd, A.J., Hammer, M.F. (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Int.* 129: 10-24.

Schoske, R., Vallone, P.M., Kline, M.C., Redman, J.W., Butler, J.M. (2004) High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Sci. Int.* 139: 107-121.

Shewale, J.G. et al. (2004) Y-chromosome STR system, Y-PLEX 12, for forensic casework: development and validation. *J. Forensic Sci.* 49: 1278-1290.

Krenke, B.E. et al. (2005) Validation of male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Sci. Int.* 151: 111-124.

Mulero, J.J. et al. (2006) Development and validation of the AmpFlSTR Yfiler™ PCR Amplification Kit: a male specific, single amplification 17 Y-STR multiplex system. *J. Forensic Sci.* 51: 64-75.

Gross, A.M. et al. (2006) Y-STR concordance study between Y-Plex5, Y-Plex6, Y-Plex12, PowerplexY, Y-Filer, MPI, and MPII. *J. Forensic Sci.* 51: 1423-1428.

Mayntz-Press, K.A. and Ballantyne, J. (2007) Performance characteristics of commercial Y-STR multiplex systems. *J. Forensic Sci.* 52: 1025-1034.

Parson, W. et al. (2008) Y-STR analysis on DNA mixture samples—results of a collaborative project of the ENFSI DNA Working Group. *Forensic Sci. Int. Genet.* 2: 238-22.

Gross, A.M. et al. (2008) Internal validation of the AmpFlSTR Yfiler amplification kit for use in forensic casework. *J. Forensic Sci.* 53: 125-134.

Additional Loci

Redd, A.J., Agellon, A.B., Kearney, V.A., Karafet, T., de Knijff, P., Park, H., Butler, J.M., Hammer, M.F. (2002) Forensic value of fourteen novel STRs on the human Y chromosome. *Forensic Sci. Int.* 130: 97-111.

Kayser, M. et al. (2004) A comprehensive survey of human Y-chromosomal microsatellites. *Am. J. Hum. Genet.* 74: 1183-1197.

Butler, J.M., Decker, A.E., Vallone, P.M., Kline, M.C. (2006) Allele frequencies for 27 Y-STR Loci with U.S. Caucasian, African American, and Hispanic Samples. *Forensic Sci. Int.* 156: 250-260.

Hanson, E.K. and Ballantyne, J. (2006) Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. *Legal Med.* 8: 110-120.

Hanson, E.K. and Ballantyne, J. (2007) An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. *PLoS ONE* 2: e688.

Decker, A.E., Kline, M.C., Vallone, P.M., Butler, J.M. (2007) The impact of additional Y-STR loci on resolving common haplotypes and closely related individuals. *FSI Genetics* 1:215-217.

Butler, J.M., Hill, C.R., Decker, A.E., Kline, M.C., Reid, T.M., Vallone, P.M. (2007) New autosomal and Y-chromosome STR loci: characterization and potential uses. *Proceedings of the Eighteenth International Symposium on Human Identification.*

Rodig, H. et al. (2008) Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci. *Forensic Sci. Int.* 174: 182-188.

Useful References on Y-Chromosome Markers and Assays

Prepared by John Butler (May 2009)

Population Variation and Data Interpretation

Kayser, M., et al. (2003) Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. *Genome Res.* 13: 624-634.

Budowle, B. et al. (2005) Twelve short tandem repeat loci Y chromosome haplotypes: Genetic analysis on populations residing in North America. *Forensic Sci. Int.* 150: 1-15.

Redd, A.J. et al. (2006) Genetic structure among 38 populations from the United States based on 11 U.S. core Y chromosome STRs. *J. Forensic Sci.* 51: 580-585.

Hammer, M.F. et al. (2006) Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. *Forensic Sci. Int* 164: 45-55.

Budowle, B. et al. (2007) Basic principles for estimating the rarity of Y-STR haplotypes derived from forensic evidence. *Proceedings of the Eighteenth International Symposium on Human Identification*. Available at http://www.promega.com/geneticidproc/

Liu, H.-M. et al. (2007) Y-chromosome short tandem repeats analysis to complement paternal lineage study: a single institutional experience in Taiwan. *Transfusion* 47: 918-926.

SWGDAM (2009) Y-Chromosome Short Tandem Repeat (Y-STR) Interpretation Guidelines. *Forensic Sci. Comm.* 11(1); available at http://www.fbi.gov/hq/lab/fsc/current/standards/2009_01_standards01.htm.

Duplication/Deletion

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.

Capelli, C. et al. (2007) Phylogenetic evidence for multiple independent duplication events at the DYS19 locus. *Forensic Sci. Int. Genet.* 1: 287-290.

Budowle, B. et al. (2008) Null allele sequence structure at the DYS448 locus and implications for profile interpretation. *Int. J. Legal Med.* 122: 421-427.

Mixture Statistics

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.

Joint Match Probability

Walsh, B. et al. (2008) Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Sci Int* 174: 234-238.

Amorim, A. (2008) A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. *Forensic Sci. Int. Genet.* 2(4): 376-378.

Mutation Rates

Kayser, M. et al. (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* 66: 1580-1588.

Dupuy, B.M. et al. (2004) Y chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. *Hum. Mutation* 23: 117-124.

Gusmão, L. et al. (2005) Mutation rates at Y chromosome specific microsatellites. *Hum. Mutation* 26: 520-528.

Decker, A.E., Kline, M.C., Redman, J.W., Reid, T.M., Butler, J.M. (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *FSI Genetics 2(3):e31-e35*.

Pollin, T.I. et al. (2008) Investigations of the Y chromosome, male founder structure and YSTR mutation rates in the Old Order Amish. *Hum. Hered.* 65: 91-104.

Ge, J. et al. (2009) Mutation rates at Y chromosome short tandem repeats in Texas populations. *Forensic Sci. Int. Genet.* 3: 179-184.

Goedbloed, M. et al. (2009) Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms including in the AmpFlSTR Yfiler PCR amplification kit. *Int. J. Legal Med.* (in press).

Useful References on Y-Chromosome Markers and Assays

Prepared by John Butler (May 2009)

Other NIST Publications

Schoske, R., Butler, J.M., Vallone, P.M., Kline, M.C., Prinz, M., Redd, A.J., Hammer, M.F. (2001) Development of Y STR megaplex assays. *Proceedings of the Twelve International Symposium on Human Identification 2001*, Promega Corporation. http://www.promega.com/geneticidproc/ussymp12proc/contents/butler.PDF

Butler, J.M. (2003) Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. *Forensic Sci. Rev.* 15:91-111.

Schoske, R., Vallone, P.M., Ruitberg, C.M., Butler, J.M. (2003) Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Anal. Bioanal. Chem.* 375: 333-343.

Butler, J.M., Schoske, R., Vallone, P.M. Highly multiplexed assays for measuring polymorphisms on the Y-chromosome. (2003) *Progress in Forensic Genetics 9* (Brinkmann, B. and Carracedo, A., eds.), Elsevier Science: Amsterdam, The Netherlands, International Congress Series 1239, pp. 301-305.

Schoske, R. (2003) The design, optimization and testing of Y chromosome short tandem repeat megaplexes. PhD dissertation, American University, 270 pp.

Vallone, P.M. and Butler, J.M. (2004) Multiplexed assays for evaluation of Y-SNP markers in U.S. populations. *Progress in Forensic Genetics 10*, Elsevier Science: Amsterdam, The Netherlands, International Congress Series 1261, 85-87.

Butler, J.M. and Schoske, R. (2004) Forensic value of the multi-copy Y-STR marker DYS464. *Progress in Forensic Genetics 10*, Elsevier Science: Amsterdam, The Netherlands, International Congress Series 1261, 278-280..

Butler, J.M. and Schoske, R. (2004) Duplication of DYS19 flanking regions in other parts of the Y chromosome. *Int. J. Legal Med.*, 118: 178-183.

Vallone, P.M. and Butler, J.M. (2004) Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *J. Forensic Sci.* 49(4): 723-732.

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*. Elsevier Academic Press: New York; Chapter 9 "Y Chromosome DNA Testing", pp. 201-239.

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.

Vallone, P.M., Fahr, K., Kostrzewa, M. (2005) Genotyping SNPs using a UV-photocleavable oligonucleotide in MALDI-TOF MS. *Methods in Molecular Biology: Forensic DNA Typing Protocols* (Carracedo, A., ed.), Humana Press: Totowa, New Jersey, 297: 169-178.

Butler, J.M. and Schoske, R. (2005) U.S. population data for the multi-copy Y-STR locus DYS464. *J. Forensic Sci.* 50(4): 975-977.

Butler, J.M., Appleby, J.E., Duewer, D.L. (2005) Locus-specific brackets for reliable typing of Y-chromosome short tandem repeat markers. *Electrophoresis* 26: 2583-2590.

Mulero, J.J., Budowle, B., Butler, J.M., Gusmão, L. (2006) Letter to the Editor--Nomenclature and allele repeat structure update for the Y-STR locus GATA H4. *J. Forensic Sci.* 51(3): 694.

Barni, F., Salata, E., Rapone, C., Kline, M., Lago, G., Butler, J., Berti, A. (2007) Characterisation of novel alleles and duplication events in the STR DYS19, DYS439, DYS389II and DYS385 loci. *Identity (Promega Newsletter for Europe, Middle East and Africa)*, Spring 2007, pp. 1-3. [.pdf]

Literature Listing prepared by John Butler (May 2009)

Low-Copy Number (LCN) and Touch DNA Evidence

Abaz, J. et al. (2002) Comparison of the variables affecting the recovery of DNA from common drinking containers. *Forensic Sci. Int.* 126: 233-240.

Alessandrini, F. et al. (2003) Fingerprints as evidence for a genetic profile: morphological study on fingerprints and analysis of exogenous and individual factors affecting DNA typing. *J. Forensic Sci.* 48: 586-592.

Budowle, B. et al. (2001) Low copy number – consideration and caution. *Proceedings of the Twelfth International Symposium on Human Identification*. See http://www.promega.com/geneticidproc/ussymp12proc/contents/budowle.pdf.

Buckleton, J. and Gill, P. (2005) Low copy number. Chapter 8 in *Forensic DNA Evidence Interpretation* (Eds. J. Buckleton, C.M. Triggs, S.J. Walsh) CRC Press: Boca Raton, FL, pp. 275-297.

Caddy, B. et al. (2008) Review of the science of low template DNA analysis. Available at http://police.homeoffice.gov.uk/publications/operational-policing/Review_of_Low_Template_DNA_1.pdf

Capelli, C. et al. (2003) "Ancient" protocols for the crime scene? Similarities and differences between forensic genetics and ancient DNA analysis. *Forensic Sci. Int.* 131: 59-64.

Evett, I. W. et al. (2002) Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *J. Forensic Sci.* 47(3): 520-530.

Findlay, I. et al. (1997) DNA fingerprinting from single cells. *Nature* 389: 555-556.

Forensic Science Regulator (2008) Response to Professor Brian Caddy's review of the science of low template DNA analysis. Available at
http://police.homeoffice.gov.uk/publications/operational-policing/response-caddy-dna-review

Gill, P. (2001) Application of low copy number DNA profiling. *Croatian Med. J.* 42(3): 229-232.

Gill, P. et al. (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci. Int.* 112(1): 17-40.

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.

Lowe, A. et al. (2002) The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Sci. Int.* 129: 25-34.

Lowe, A. et al. (2003) Use of low copy number DNA in forensic inference. *Progress in Forensic Genetics 9 – International Congress Series* 1239: 799-801.

Murray, C. et al. (2001) Use of low copy number (LCN) DNA in forensic inference. *Proceedings of the Twelfth International Symposium on Human Identification*. Available at http://www.promega.com/geneticidproc/.

Polley, D. et al. (2006) An investigation of DNA recovery from firearms and cartridge cases. *Canadian Soc. Forensic Sci. J.* 39: 217-228.

Raymond, J.J. et al. (2008) Trace DNA analysis: do you know what your neighbour is doing? A multi-jurisdictional survey. *Forensic Sci. Int. Genet.* 2: 19-28.

Rutty, G.N. et al. (2003) The effectiveness of protective clothing in the reduction of potential DNA contamination of the scene of crime. *Int. J. Legal Med.* 117: 170-174.

Schiffner, L.A. et al. (2005) Optimization of a simple, automatable extraction method to recover sufficient DNA from low copy number DNA samples for generation of short tandem repeat profiles. *Croatian Med. J*. 46(4): 578-586.

Taberlet, P. et al. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res*. 24: 3189-3194.

Van Hoofstat, D.E.O. et al. (1998) DNA typing of fingerprints and skin debris: sensitivity of capillary electrophoresis in forensic applications using multiplex PCR. *Proceedings of the Second European Symposium on Human Identification*, pp. 131-137. Madison, Wisconsin: Promega Corporation. Available at http://www.promega.com/geneticidproc/eusymp2proc/

Van Oorschot, R.A. and Jones, M. (1997) DNA fingerprints from fingerprints. *Nature* 387: 767.

Walsh, P.S. et al. (1992) Preferential PCR amplification of alleles: Mechanisms and solutions. *PCR Meth. Appl.* 1: 241-250.

Whitaker, J.P. et al. (2001) A comparison of the characteristics of profiles produced with the AMPFISTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci. Int.* 123(2-3): 215-223.

Wickenheiser, R.A. (2002) Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. *J. Forensic Sci.* 47(3): 442-450.

Mixtures

Torres, Y. et al. (2003) DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic Sci. Int.* 134: 180-186.

RECOMMENDATIONS

Gill, P. et al. (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. *Forensic Sci. Int. Genet.* 2: 76-82.

Morling, N. et al. (2007) Interpretation of DNA mixtures—European consensus on principles. *Forensic Sci. Int. Genet.* 1: 291-292.

Schneider, P.M. et al. (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *Int. J. Legal Med.* 123: 1-5.

STRATEGY

Buckleton, J.S. et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *Forensic Sci. Int. Genet.* 1:20-28.

Clayton, T.M. et al. (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.* 91: 55-70.

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

Curran, J.M. (2008) A MCMC method for resolving two person mixtures. *Sci. Justice* 48: 168-177.

Gill, P. et al. (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.

Gill, P. et al. (2008) Interpretation of complex DNA profiles using empirical models and a method to measure their robustness. *Forensic Sci. Int. Genet.* 2: 91-103.

Leclair, B. et al. (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.

Paoletti, D.R. et al. (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. *J. Forensic Sci.* 50(6): 1361-1366.

Wickenheiser, R. (2006) General guidelines for categorization and interpretation of mixed STR DNA profiles. *Canadian Soc. Forensic Sci. J.* 39: 179-216.

SOFTWARE

Bill, M. et al. (2005) PENDULUM-a guideline-based approach to the interpretation of STR mixtures. *Forensic Sci. Int*. 148: 181-189.

FSS-i3 software (from FSS and Promega):
http://www.promega.com/fssi3
http://www.forensic.gov.uk/forensic_t/i3/index.htm

GeneMapperID-X (from Applied Biosystems):
http://idx.appliedbiosystems.com

Oldroyd, N. and Schade, L.L. (2008) Expert assistant software enables forensic DNA analysts to confidently process more samples. *Forensic Magazine* 5(6): 25-28.

TrueAllele (from Cybergenetics):
http://www.cybgen.com/

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

STATISTICS

Buckleton, J. and Curran, J. (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci. Int. Genet.* 2(4): 343-348.

Curran, J.M. et al. (1999) Interpreting DNA mixtures in structured populations. *J. Forensic Sci.* 44(5): 987-995.

DNA Advisory Board (2000) Statistical and population genetic issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). *Forensic Sci. Comm*. 2(3); http://www.fbi.gov/programs/lab/fsc/backissu/july2000/dnastat.htm.

Ladd, C. et al. (2001) Interpretation of complex forensic DNA mixtures. *Croatian Med. J.* 42(3): 244-246.

Weir, B.S. et al. (1997) Interpreting DNA mixtures. *J. Forensic Sci.* 42(2): 213-222.

INTERLABORATORY STUDIES

Duewer, D.L. et al. (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. et al. (2004) NIST mixed stain study 3: signal intensity balance in commercial short tandem repeat multiplexes. *Anal. Chem.* 76: 6928-6934.

Kline, M.C. et al. (2003) NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Anal. Chem.* 75: 2463-2469.

NIST Mixture Interpretation Study 2005 (MIX05): http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm

Int J Legal Med (2009) 123:1–5
DOI 10.1007/s00414-008-0244-4

REVIEW ARTICLE

The German Stain Commission: recommendations for the interpretation of mixed stains

P. M. Schneider · R. Fimmers · W. Keil · G. Molsberger · D. Patzelt · W. Pflug · T. Rothämel · H. Schmitter · H. Schneider · B. Brinkmann

Received: 8 April 2008 / Accepted: 16 April 2008 / Published online: 5 July 2008

Abstract In the course of forensic DNA analysis, the interpretation of DNA profiles of mixed stains, i.e. cell material from more than a single donor, has become increasingly more important. The German Stain Commission, a joint commission of Institutes of Forensic Science and Legal Medicine, has therefore developed guidelines aiming to harmonize the evaluation of mixed stains in German criminal cases.

Keywords Short tandem repeat typing · Biostatistical analysis · Likelihood ratio · Probability of exclusion · Mixtures

Preface

Since the beginning of forensic stain analysis, mixed stains have been observed [1, 2]. Over the past few years, they have

Schneider, Keil, Molsberger, Patzelt, Pflug, Rothämel, Schmitter, Schneider, and Brinkmann are members of the Stain Commission, the joint commission of Institutes of Legal Medicine and Forensic Science in Germany.

P. M. Schneider
Institut für Rechtsmedizin, Universitätsklinikum Köln,
Melatengürtel 60-62,
50823 Cologne, Germany

W. Keil
Institut für Rechtsmedizin, Universität München,
Nußbaumstraße 26,
80336 Munich, Germany

H. Schmitter
Bundeskriminalamt,
Äppelallee 45,
65203 Wiesbaden, Germany

R. Fimmers
Institut für Medizinische Biometrie,
Informatik und Epidemiologie, Universität Bonn,
Sigmund-Freud-Str. 25,
53105 Bonn, Germany

B. Brinkmann (✉)
Forensische Genetik,
Röntgenstr. 23,
48149 Münster, Germany
e-mail: brinkma@uni-muenster.de

G. Molsberger
Landeskriminalamt Nordrhein-Westfalen,
Völklingerstr. 49,
40221 Düsseldorf, Germany

D. Patzelt
Institut für Rechtsmedizin,
Julius-Maximilians-Universität,
Versbacher Str. 3,
97078 Würzburg, Germany

W. Pflug
Landeskriminalamt Baden-Württemberg,
Taubenheimstr. 85,
70372 Stuttgart, Germany

T. Rothämel
Institut für Rechtsmedizin, Medizinische Hochschule,
Carl-Neuberg-Str. 1,
30625 Hannover, Germany

H. Schneider
Hessisches Landeskriminalamt,
Hölderlinstrasse 5,
65187 Wiesbaden, Germany

gained importance as evidence due to improved analytical methods and the enormous increase in the numbers of investigated stains [3, 4]. While the interpretation of single source stains usually does not cause problems [5], the evaluation and interpretation of mixed DNA stains requires particular attention [6–8]. Our recommendations – first published in German [9] – are intended to build a framework for an adequate means of treating typical cases. However, it is beyond the scope of these basic recommendations to address all possible constellations.

Definitions

A stain exhibiting more than two alleles in a single DNA system[1] shall be considered a mixed stain except in the case of genetic irregularities (e.g., trisomy, somatic mosaicism, or duplication). If more than two alleles are observed in at least two DNA systems, the presence of a mixed stain shall be assumed.

The number of possible contributors to a mixed stain shall be derived, if possible:

- In general, the presence of not more than four alleles in a given system allows the assumption of at least two independent stain donors.
- In general, the presence of not more than six alleles in a given system allows the assumption of at least three independent stain donors.
- In general, if more than six alleles are observed in a given system, the exact number of stain donors cannot be reliably determined.

Classification of mixed stains

Type A has no obvious major contributor with no evidence of stochastic effects.[2] Type B has clearly distinguishable major and minor DNA components; consistent peak height ratios of approximately 4:1 (major to minor component) across all heterozygous systems, and no evidence of stochastic effects. Type C has mixtures with no major component(s) and evidence of stochastic effects.

[1] A DNA system is a genetic locus exhibiting a short tandem repeat polymorphism amplified with a pair of defined primers using the polymerase chain reaction (PCR).

[2] DNA profiles obtained from the amplification of samples with low DNA content and/or poor DNA quality, where the occurrence of allelic drop out and/or locus drop out has to be assumed.

Evaluation criteria

Peak analysis

The morphology of a peak shall be typical and fully consistent with an allele of a given short tandem repeat system. Generally, reproducible peaks with heights >50 relative fluorescence units (RFU) can be considered regular peaks if the noise of the baseline is low and the number of PCR cycles recommended by the manufacturer was used.

The presence of peaks exhibiting a low signal strength (i.e., typically below 100 RFU) and/or peaks exhibiting clearly variable intensities shall be annotated in the table of observed alleles. Tables in the final report shall be accompanied by a legend explaining the designations of peak characteristics.

Stutter peaks

Both $n-1$ and $n+1$ stutter peaks may occur. Their heights depend on the DNA systems and the amplification conditions. A stutter peak may, in certain cases, exhibit up to 15% of the height of the corresponding main peak. Furthermore, the following shall be considered for the evaluation of a stutter peak:

- The relative stutter intensities of the alleles of a locus, as well as those between loci of a multiplex amplification.
- The possibility that a stain allele is in the position of a stutter peak.

In case of reasonable doubt, a peak in the position of a stutter peak shall be considered a true allele and part of the DNA profile and shall be included in the biostatistical calculation.

Inclusion/exclusion criteria

Inclusion

If all alleles of a person in question are uniformly present in a mixed stain, the person shall be considered a possible contributor to the stain.

Exclusion

If alleles of a person in question are not present in a mixed stain, the person shall not be considered as a possible contributor to the stain.

Grey area between inclusion and exclusion

The following effects may occur in type C mixtures due to imbalances between the mixture components and may cause

difficulties in reaching an unambiguous decision about inclusion or exclusion across all analyzed DNA systems:

- Locus drop out and allelic drop out (e.g., caused by the sensitivity of the amplification system, as well as by stochastic effects).
- Allelic drop out is more likely to occur for longer than for shorter alleles, and in particular for DNA systems with long amplicon sizes.

Additional criteria

In every case, the decision about inclusion or exclusion shall be made after careful consideration of the issues described under the "Grey area between inclusion and exclusion" section. The reasons shall be explained in detail. If appropriate, it shall be stated why a clear decision about inclusion or exclusion was not possible.

Biostatistical calculations for mixed stains

Basis

The basis for all calculations is the knowledge of the allele frequencies in the relevant population.

Probability of exclusion (P_E)/probability of inclusion (P_I)

P_I represents the combined probability (relative population frequency) of all combinations of genotypes that cannot be excluded to have contributed to the DNA profile of a stain based on the criteria given in the "Inclusion" section. P_I is equivalent to the match probability in the case of a stain originating from a single person.

The calculation of P_I is independent of assumptions about the number of possible contributors to a stain, the genotypes, and the ethnic origin of persons involved in a given case. It is equivalent to the probability that a randomly selected person is a contributor to the stain [=random man not excluded (RMNE)]. The probability of exclusion $P_E = 1 - P_I$ indicates the probability of excluding a randomly selected person as a contributor to a given stain.

Likelihood ratio

The calculation of the likelihood ratio (LR) is based on the assumption of two mutually excluding hypotheses. This imperatively requires the description of a distinct scenario for a given stain case. Both hypotheses explicitly describe alternative scenarios for the origin of a stain. Each of these hypotheses shall clearly state who contributed to the stain and how many unknown contributors are assumed. Then, a calculation of the likelihood for the occurrence of the DNA profile of the stain is performed based on the assumption of the respective hypotheses: $L(\text{stain}|H)$. The LR

$$\text{LR} = \frac{L(\text{stain}|H_1)}{L(\text{stain}|H_2)}$$

allows the evidential value of a stain to be calculated with reference to a specific person involved in a case, e.g., an accused stain donor.

Given a two-person mixed stain M and that all observed alleles can be explained by the genotype of the victim, G_v, and the genotype of the suspect, G_s, the hypotheses can be formulated as follows:

Hypothesis H_p (view of the prosecution): The stain M originates from the victim V and the suspect S.

Hypothesis H_d (view of the defense): The stain originates from the victim V and from an unknown person U unrelated to the suspect.

$$\text{LR} = \frac{L(M|H_p)}{L(M|H_d)} = \frac{L(M|G_v, G_s)}{L(M|G_v, G_u)}$$

The resulting LR provides a numerical value, which indicates how many times more likely the observed DNA profile is under the assumption of the scenario described in H_p compared to the scenario described in H_d.

Procedures

Calculation for a mixed stain with an unambiguous major component from one person

The conclusion of a major DNA profile from a single contributor in a mixed stain shall only be drawn if a peak height ratio of at least 4:1 (major vs minor component) is observed across all heterozygous DNA systems (see "Definitions" section). In this case, the major DNA profile can be considered equivalent to that of a stain originating from a single person, and all calculations can be performed accordingly.

Calculation based on the LR

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, e.g., the suspect, is part of the minor component of the mixture (see "Definitions" section)]

then the calculation of a LR is appropriate.

Calculation based on probability of exclusion/inclusion

If a major DNA profile cannot be identified based on unambiguous DNA profiles, or if the number of contributors cannot be determined, calculations of the probability of exclusion P_E or the probability of inclusion P_I, respectively, for randomly selected persons is appropriate. Also, the calculation of P_E and P_I is always possible for type A and type B mixtures.

Supplementary recommendations

Further calculations that may result in erroneous interpretations of the evidence shall not be performed (e.g. reporting the genotype frequency of a non-excluded suspect, if the mixed stain does not allow a meaningful biostatistical interpretation).

Validated computer programmes for the calculation of complex mixed stains are available.

Acknowledgements We thank K. Anslinger (Munich), P. Berschick (Düsseldorf), M. Eckert (Wiesbaden), C. Hohoff (Münster), S. Jung (Würzburg), and J. Schnee-Griese (Stuttgart) for helpful collaboration. Furthermore, we thank N. Morling (Copenhagen) and S. Rand (Antwerp) for a critical revision of the English text.

Appendix

Examples of the calculations of P_I and P_E

The probability of inclusion P_I is calculated from the sum of all genotypes of possible stain contributors. In a stain case, where a, b, and c denote the alleles of a DNA system detected in the mixture, the sum of all relevant genotypes can be calculated as follows (assuming that allele frequency data conform to Hardy–Weinberg equilibrium):

$$P_I = a^2 + b^2 + c^2 + 2ab + 2bc + 2ac$$

This term can be simplified using the formula for the binominal distribution:

$$a^2 + b^2 + c^2 + 2ab + 2bc + 2ac = (a + b + c)^2$$

Assuming a frequency of 0.1 for alleles a, b, and c, the following result is obtained:

$$P_I = 0.3^2 = 0.09$$

Thus, it is expected that 9% of a group of randomly selected persons will not be excluded as stain contributors. This is equivalent to one out of 11 randomly selected persons (=RMNE). The probability of exclusion is calculated from the difference

$$P_E = 1 - P_I = 1 - 0.09 = 0.91$$

Thus, it is expected that 91% of a group of randomly selected persons will be excluded as stain contributors. For several DNA systems, S_1, S_2,…, S_n, which are genetically unlinked (i.e., in linkage equilibrium), the general expression of $P_E(S_1, S_2,\ldots, S_n)$ can be derived from the product of the individual inclusion probabilities $P(S_j)$ as follows:

$$P_E(S_1, S_2, \ldots, S_n) = 1 - [P_I(S_1) \cdot P_I(S_2) \cdot \ldots \cdot P_I(S_n)]$$

Examples for the calculation of the LR

Simple scenario

Consider a case with a mixed stain M with three alleles, a, b, and c, composed from a victim and a perpetrator. The victim V has the genotype AB, and the suspect S has the genotype BC. The hypotheses can be given as follows:

> H_p: The stain M originates from the victim V and the suspect S.
> H_d: The stain M originates from the victim V and from an unknown person unrelated to the suspect.

Let us first derive the numerator of the LR. The prosecution claims that the stain can be explained by a combination of the genotypes of the victim and the suspect, as there are no unaccounted alleles. Hence, the numerator results as

$$L(M|H_p) = L(M|G_v, G_s) = 1$$

The defense, however, claims that the suspect has not contributed to the stain. The genotype of the suspect is not relevant since the presence of allele c in the mixture must be explained by the contribution of an unknown person. As allele c may have been contributed either by a person homozygous for allele c or from a person heterozygous for c in combination with allele a or b, the denominator is as follows:

$$L(M|H_d) = L(M|G_v, G_u) = 2ac + 2bc + c^2$$

And, thus, the entire expression is given as

$$\mathrm{LR} = \frac{1}{2ac + 2ab + c^2}$$

Assuming a frequency of 0.1 for alleles a, b, and c, the following result is obtained:

$$\mathrm{LR} = \frac{1}{0.02 + 0.02 + 0.01} = \frac{1}{0.05} = 20$$

The result can be described by the following statement: It is 20 times more likely to observe the DNA profile if the mixed stain originated from the victim and the suspect than if it originated from the victim and an unknown person (who is unrelated to the suspect[3]).

Complex scenario

Let us consider a case with a mixed stain M with four alleles a, b, c, and d found on the victim's clothes. The victim's genotype is EF and, hence, the corresponding alleles e and f are not observed in the stain. Suspect S has genotype AB, but there is no known second person who may have contributed the alleles c and d. The hypotheses can be given as follows:

H_p: Stain M originates from suspect S and an unknown person U.
H_d: Stain M originates from two unknown persons $U1$ and $U2$.

The prosecution claims that the stain can be explained by a combination of the suspect's genotype and a second person with the genotype CD. Hence, the numerator results as

$$L(M|H_p) = L(M|G_s, G_u) = 2cd$$

The defense claims that no genotypes of the contributors are known. Thus, the sum of all possible genotype combinations from two persons $U1$ and $U2$ must be considered for the denominator:

Genotypes $U1$	$U2$	Combined frequency $U2$
AB	CD	$2ab \times 2cd = 4abcd$
AC	BD	$4abcd$
AD	BC	$4abcd$
BC	AD	$4abcd$
BD	AC	$4abcd$
CD	AB	$4abcd$
$L(M\|H_d) = L(M\|G_{U1}, G_{U2}) =$		$24abcd$

After reducing the term and by assuming a frequency of 0.1 for alleles a, b, c, and d, the following result is obtained:

$$\mathrm{LR} = \frac{2cd}{24abcd} = \frac{1}{12ab} = \frac{1}{0.12} = 8.3$$

[3] A familial relationship between S and the unknown stain contributor can be considered for calculating LR. However, the exact degree of relationship must be known.

Thus, it is eight times more likely to observe the DNA profile if the mixed stain originated from the suspect and an unknown person than if it originated from two unknown persons. If two suspects $S1$ and $S2$ with the genotypes AB and CD are considered for the same mixed stain scenario, the hypotheses and, hence, the LR change, as no unknown person remains for H_p:

H_p: Stain M originates from the suspects $S1$ and $S2$.
H_d: Stain M originates from two unknown persons $U1$ and $U2$.

Thus, the numerator of the LR is, again, 1. The term cannot be reduced further and the resulting LR is as follows:

$$\mathrm{LR} = \frac{1}{24abcd} = \frac{1}{0.0024} = 416.7$$

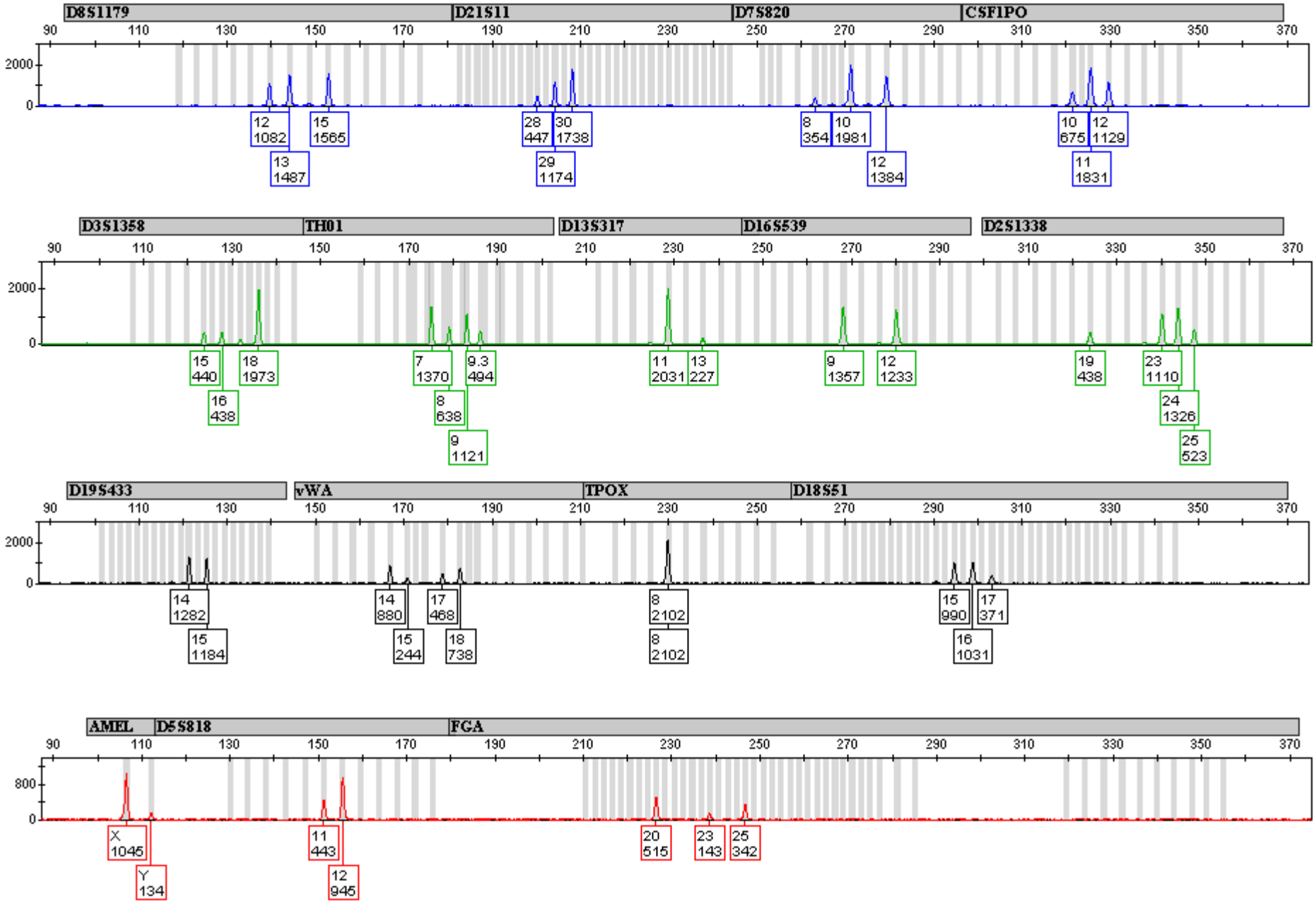
Thus, it is 416 times more likely to observe the DNA profile if the mixed stain originated from suspects $S1$ and $S2$ than if it originated from two unknown persons.

We give the following caveat: Additional hypotheses, which are not discussed here, can be formulated. Depending on the precise scenario, such additional hypotheses may be highly relevant in a given case, such as (a) H_p: the stain originates from $S1$ and $S2$; H_d: the stain originates from $B1$ and U, or (b) H_p: the stain originates from $S1$ and $S2$; H_d: the stain originates from $S2$ and U. Depending on the genotype frequencies of $S1$ and $S2$, the resulting LRs may differ significantly.

References

1. Wiegand P, Rolf B (2003) Analyse biologischer Spuren, Teil 1. Rechtsmedizin 13:103–113
2. Hohoff C, Brinkmann B (2003) Trends in der forensischen Molekulargenetik. Rechtsmedizin 13:183–189
3. Pflug W, Nguyen TMH, Merkel J (1997) Zuordnung von Schußwaffen mittels DNA-Analyse. Kriminalistik 12:799–802
4. Schöneberg A, Gerl L, Oesterreich W, Bastisch I, Gerhard M, Kärgel HJ, Fesefeldt A, Pflug W (2003) DNA-Analyse von Hautabriebspuren. Kriminalistik 8–9:497–499
5. National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington, D.C.
6. Evett IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic science. Sinauer, Sunderland
7. Buckleton J, Triggs CM, Walsh SJ (2005) Forensic DNA evidence interpretation. CRC, London
8. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): recommendations on the interpretation of mixtures. Forensic Sci Int 160:90–101
9. Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2006) Allgemeine Empfehlungen der Spurenkommission zur Bewertung von DNA-Mischspuren. Rechtsmedizin 16:401–404

Mixture Homework



Data courtesy of Amy Christen, Marshall University NEST Project Team