17th Annual National CODIS Conference
(Jacksonville, FL) – November 14, 2011

NIST Update

John M. Butler

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

Presentation Topics

- **Group Research Overview**
  - New Standard Reference Material SRM 2391c
  - STRBase website updates
  - *Advanced Topics in Forensic DNA Typing: Methodology*
  - Insertion/Deletion markers
  - STR kit concordance testing
  - Variant allele sequencing
  - Mixture interpretation training & TrueAllele evaluation
  - Rapid DNA
  - ABI 3500 validation

- **Additional Support to the Forensic DNA Community**
  - Update on ABI 3500 open letter status
  - Response to public criticism of expanded CODIS core loci

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 **a non-regulatory agency within 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.**

$686 for 3 jars

DNA typing standard

NIST
Applied Genetics

NIST Applied Genetics Group

Group Leader

John Butler

Marcia Holden

Margaret Kline

Pete Vallone

Mike Coble

Ross Haynes

Becky Hill

Erica Butts

Kristen O'Connor

Kevin Kiesler

NIST Applied Genetics

Our FY2011 Group Productivity

(Oct 2010 to Sept 2011)

- **21 publications**
  - 20 articles + 1 book
- **77 presentations**
  - 65 talks (58 invited) + 12 posters (all available on STRBase)
- **10 training workshops**
  - Mixture interpretation (ISHI, AAFS, NFSTC, IN, HI, AZ, MI, Palm Beach, Houston)
  - Capillary electrophoresis (ISFG)
- **3 Standard Reference Materials (SRMs) completed**
  - 2391c (forensic STRs), 2393 (HD), 2366 (CMV)
- **10 committee assignments**
  - VA SAC, DOD DNA oversight, FBI new CODIS core loci, SWGDAM (mixture interpretation, rapid DNA, enhanced detection methods), NIST/NIJ evidence preservation TWG, JCTLM, NIJ DNA TWG, ATCC cell line authentication

NIST Human Identity Project Teams
within the Applied Genetics Group

Forensic DNA Team

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

John Butler

STRBase, Workshops & Textbooks

Mike Coble

Mixtures, mtDNA & Y

Becky Hill

Concordance & LT-DNA

Margaret Kline

SRM work, variant alleles & Cell Line ID

Guest Researcher

Manuel **Fonde**vila Alvarez

Data Analysis Support

Dave Duewer

DNA Biometrics Team

Funding from the **FBI S&T Branch** through NIST Information Access Division

Pete Vallone

Rapid PCR, Direct PCR & Biometrics

Erica Butts

ABI 3500 & DNA Extraction

Kevin Kiesler

PLEX-ID & NGS Exploration

Office Manager
Patti Rohmiller

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

NIST STRBase Website

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

John Butler

Forensic STR Information

- STRs101: Brief Introduction to STRs
- ***Core Loci***: FBI CODIS Core STR Loci and European Core Loci
- STR Fact Sheets (observed alleles and PCR product sizes)
- **Multiplex STR kits**
- Sequence Information (annotated)
- Variant Allele Reports
- Tri-Allelic Patterns
- Mutation Rates for Common Loci
- Published PCR primers
- Y-chromosome STRs
- Low-template DNA Information *Updated*
- Mixture Interpretation *NEW*
- Kinship Analysis *NEW*
- miniSTRs (short amplicons)
- Null Alleles - discordance observed between STR kits
- STR Reference List - *now 3400 references*

Cataloged as of Oct 2011

593 variant alleles

239 tri-allelic patterns

We invite labs to supply information on variant and tri-alleles observed

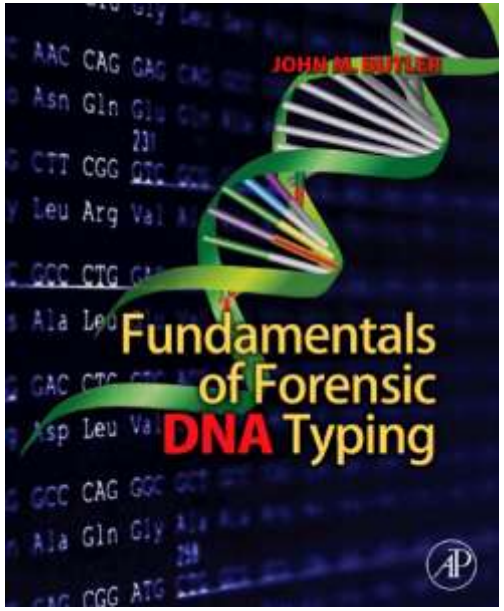
Forensic DNA Typing Textbook

3rd Edition is Three Volumes

Now part of my job at NIST (no royalties are received)

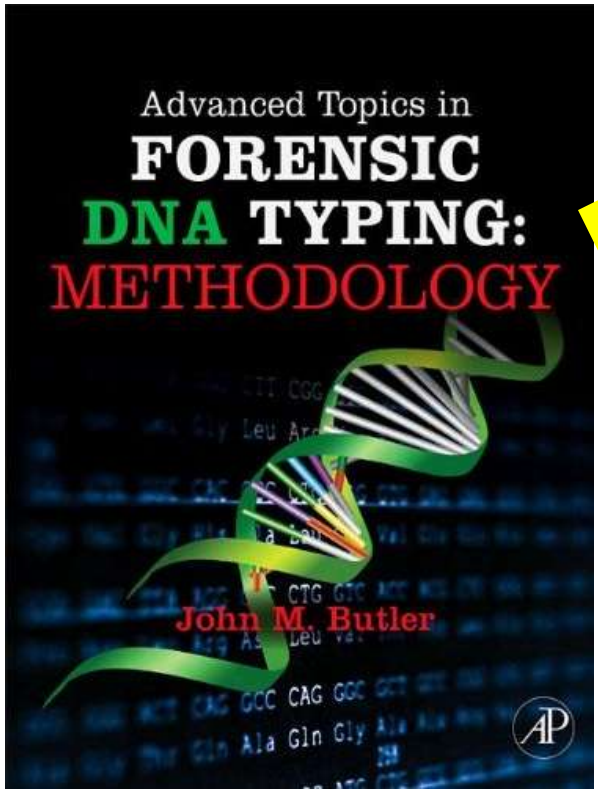
John Butler

For beginning students, general public, & lawyers



Sept 2009

~500 pages



August 2011

~700 pages

Currently being written

Advanced Topics in Forensic DNA Typing: **INTERPRETATION**

Fall 2012

~500 pages

Current NIST Projects

Short Overviews…

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

NIST SRM 2391c


Margaret Kline


Becky Hill

Main Points:

- Traceable physical reference materials to ensure accurate and comparable measurements between laboratories
- Helps meet ISO 17025 needs for traceability to a national metrology institute
- **http://www.nist.gov/srm**
- **SRM 2391c released Aug 2011**

http://www.promega.com/resources/articles/profiles-in-dna

The Latest and Greatest NIST PCR-Based DNA Profiling Standard: Updates and Status of...

The Latest and Greatest NIST PCR-Based DNA Profiling Standard: Updates and Status of Standard Reference Material® (SRM) 2391c

Article | Figures & Tables | Share

Margaret C. Kline, Carolyn R. (Becky) Hill, Jamie L. Almeida, Erica L.R. Butts, Michael D. Coble and John M. Butler

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

2011

Presentations/Publications:

- *Profiles in DNA* article (Sept 2011)
- ISFG 2011 and ISHI 2011 posters
- Forensic Sci. Int. Genet. Suppl. Ser. (2011)

NIST Standard Reference Material (SRM) for Forensic DNA Testing

SRM 2391b (2003-2011)

- **48 autosomal STR loci** with certified values
- **10 liquid genomic DNA** components + **2 punches** (cells on 903 paper)
- All single source samples
- 4 males + 6 females
- 9947A & 9948 included

SRM 2391c (2011-future)

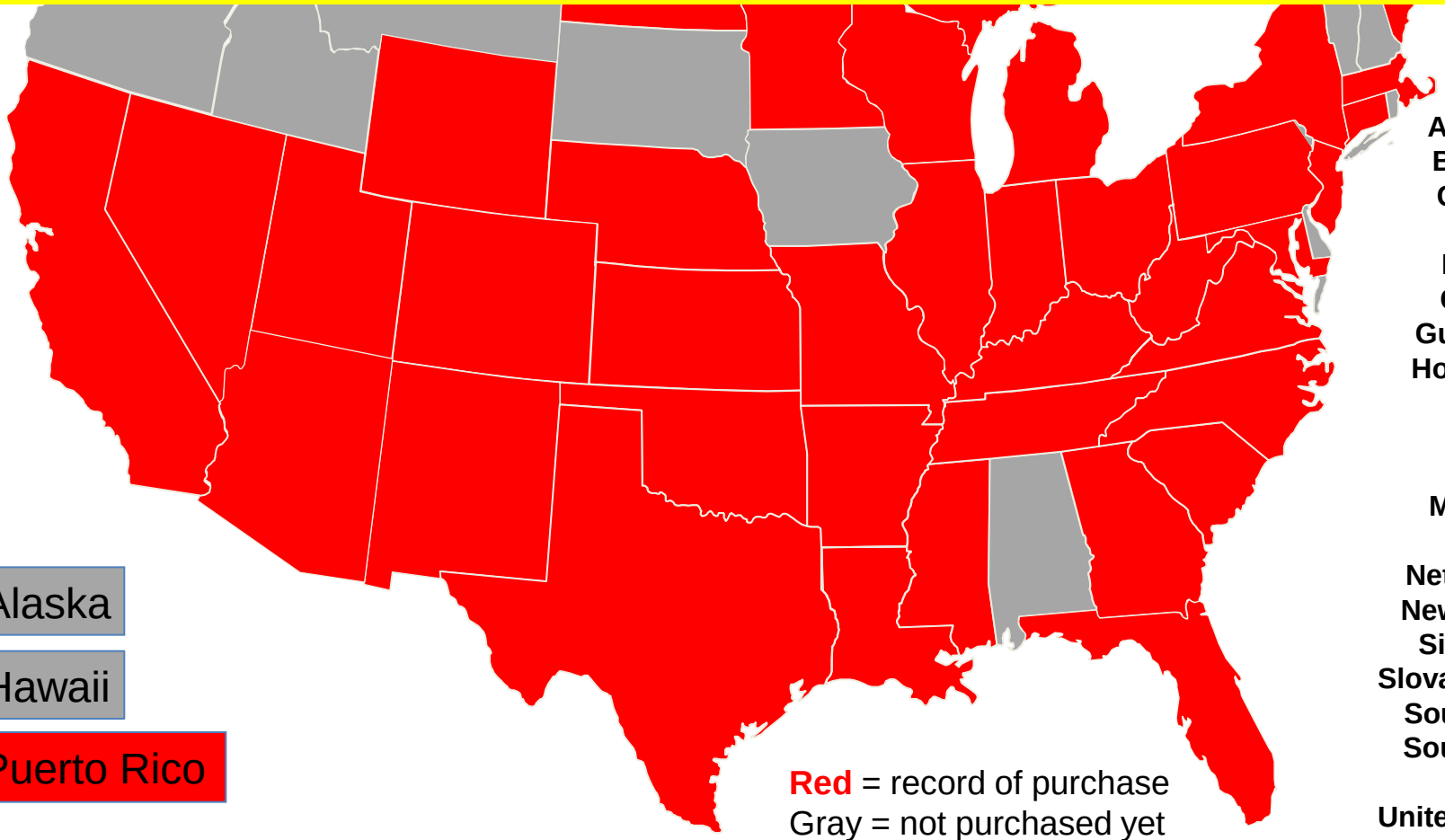
- **23 autosomal STR loci and 17 Y-STRs** certified
- **4 liquid genomic DNA** components + **2 punches** (cells on **FTA** & 903 paper)
- 5 single source + 1 mixture
- 3 males + 2 females (unique)
- All new samples
  - no 9947A or 9948

SRM 2391c to replace SRM 2391b and SRM 2395 (for Y-STRs)

Sources of NIST SRM 2395 Purchases (2004-2011)

38 states and 22 other countries

There is approximately a year's supply of SRM 2395 left for Y-STR testing – **currently the plan is to NOT replace it** and to certify the components of SRM 2391c for Y-STRs used by the forensic DNA community.



Alaska

Hawaii

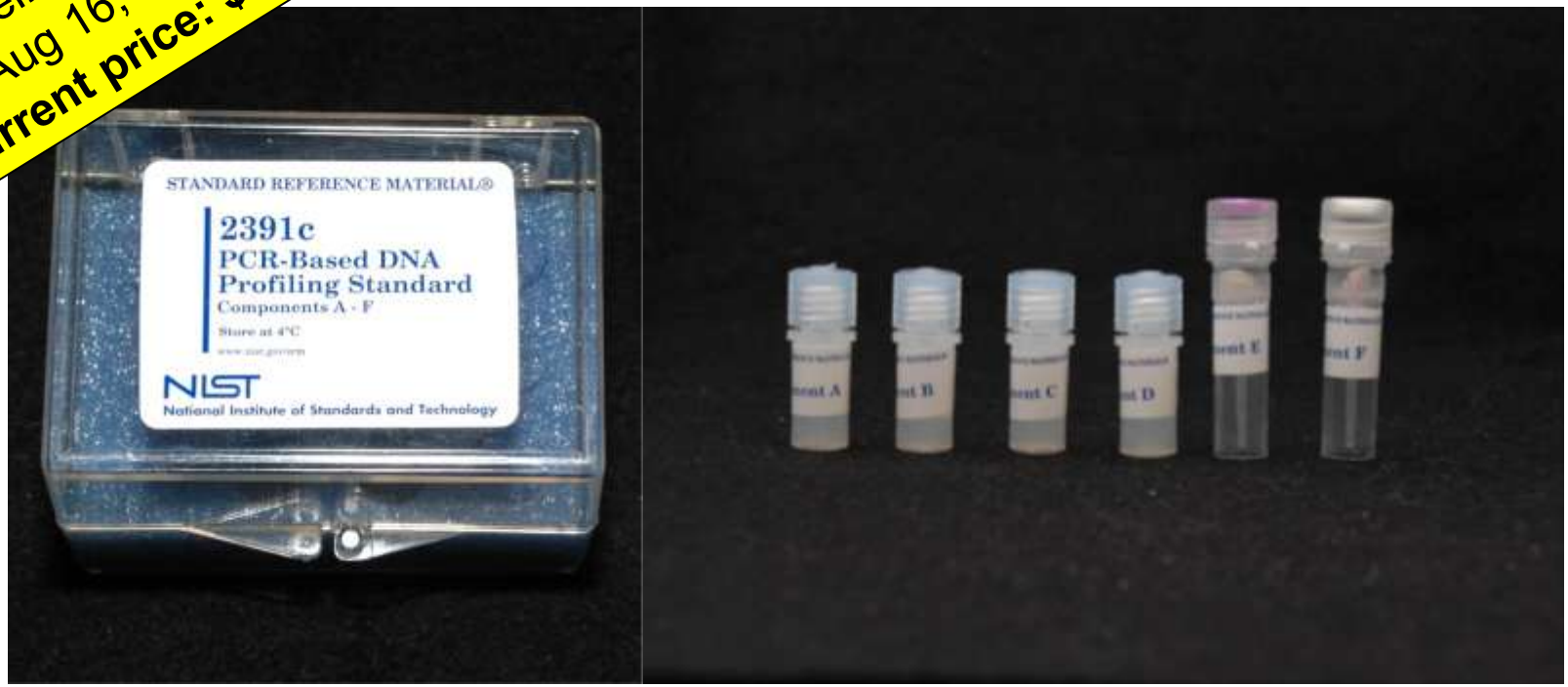
Puerto Rico

Red = record of purchase
Gray = not purchased yet

Australia
Belgium
Canada
China
France
Greece
Guatemala
Hong Kong
India
Italy
Japan
Malaysia
Malta
Netherlands
New Zealand
Singapore
Slovak Republic
South Africa
South Korea
Spain
United Kingdom
West Indies

Selling since
Aug 16, 2011
Current price: $614

NIST SRM 2391c



Produced with an entirely new set of genomic DNA samples.

9947A & 9948 are NOT included.

https://www-s.nist.gov/srmors/view_detail.cfm?srm=2391C

Description of Components in SRM 2391c

Component	Description	Quantity [a]
A	50 µL of anonymous **female** genomic DNA	1.4 – 1.9 ng DNA/µL
B	50 µL of anonymous **male** genomic DNA	1.3 – 1.5 ng DNA/µL
C	50 µL of anonymous **male** genomic DNA	1.3 – 2.0 ng DNA/µL
D	50 µL of **mixed-source** (Components A and C)	1.4 – 2.0 ng DNA/µL
E	Two 6 mm punches of CRL-1486 cells spotted on **903 paper**	~75,000 cells per punch
F	Two 6 mm punches of HTB-157 cells spotted on **FTA paper**	~75,000 cells per punch

[a] DNA concentrations and cell counts are nominal values and are **not** intended for use as quantitative standards.

STR Genotyping kits and primer mixes used at NIST to certify SRM 2391c

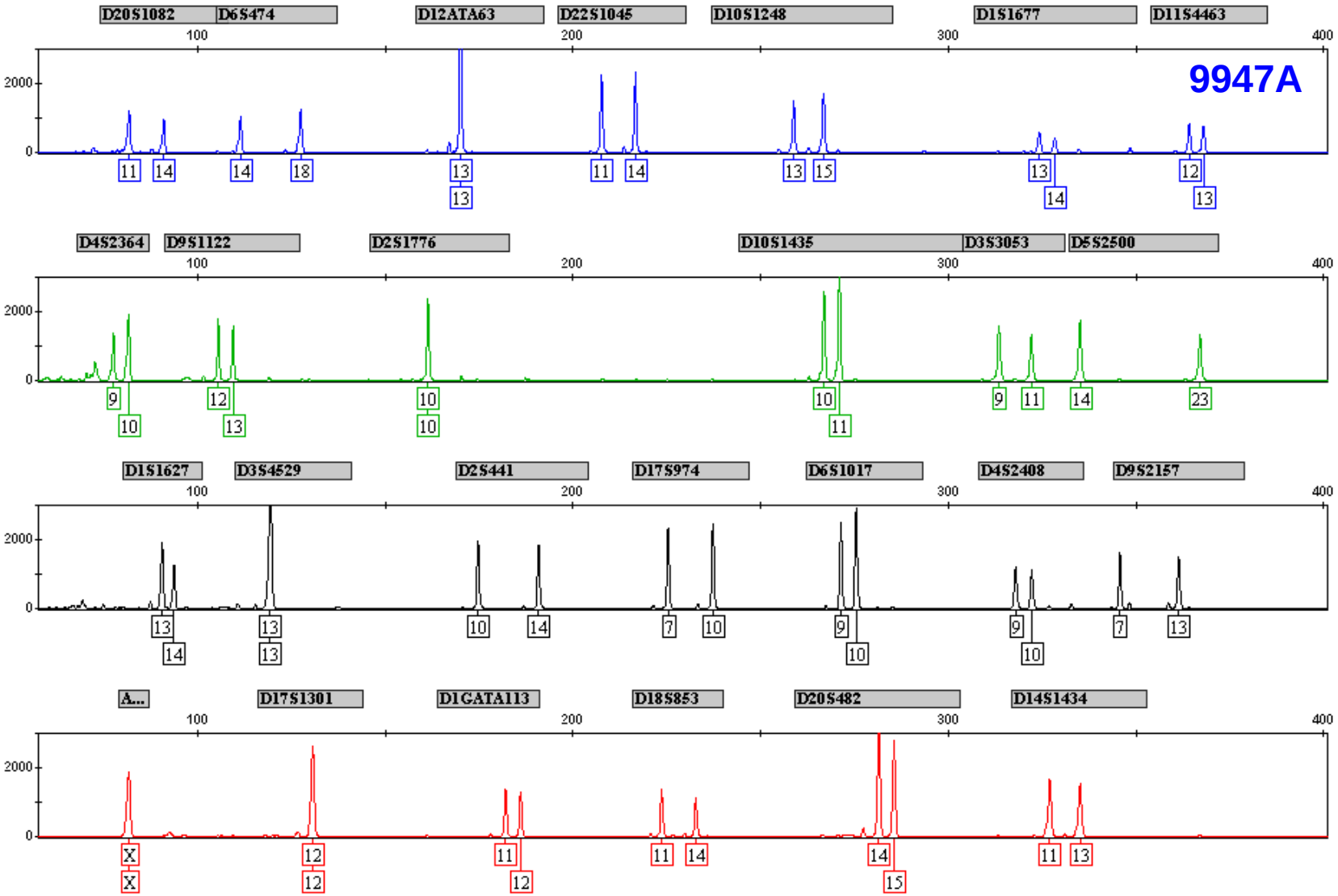
Kit Provider			Primer Mixes
Life Technologies	*Promega*	*Qiagen*	*NIST*
Identifiler	Powerplex 16	ESSplex	26plex
Identifiler Plus	Powerplex 16 HS	IDplex	miniSTRs
NGM	Powerplex ESX 17		
NGM SElect	Powerplex ESI 17		
COfiler	Powerplex ES		
Profiler	Powerplex S5		
Profiler Plus	**Powerplex Y**		
Profiler Plus ID	FFFL		
SGM Plus			
SEfiler			
MiniFiler			
Yfiler			

All results are concordant across all kits.

In total there is data for 51 autosomal STRs and 17 Y-STRs

NIST STR 26plex

Hill et al. (2009) *Journal of Forensic Sciences*, 54(5):1008-1015

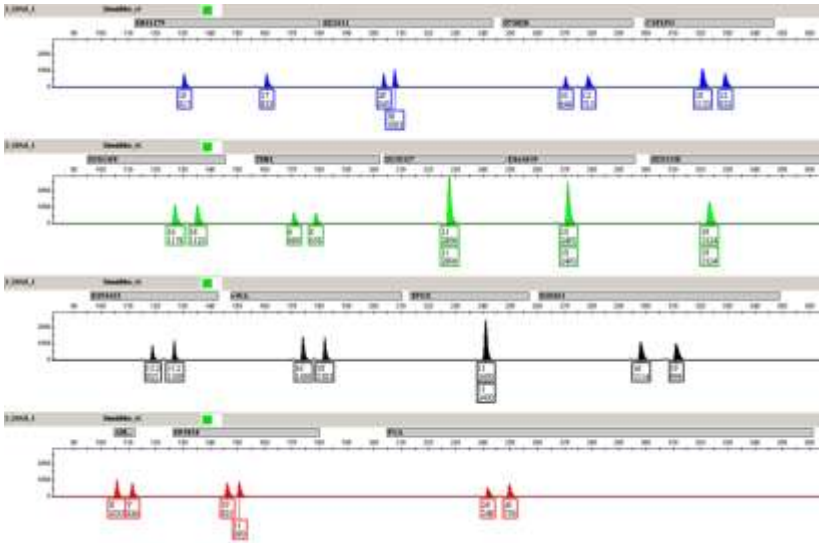


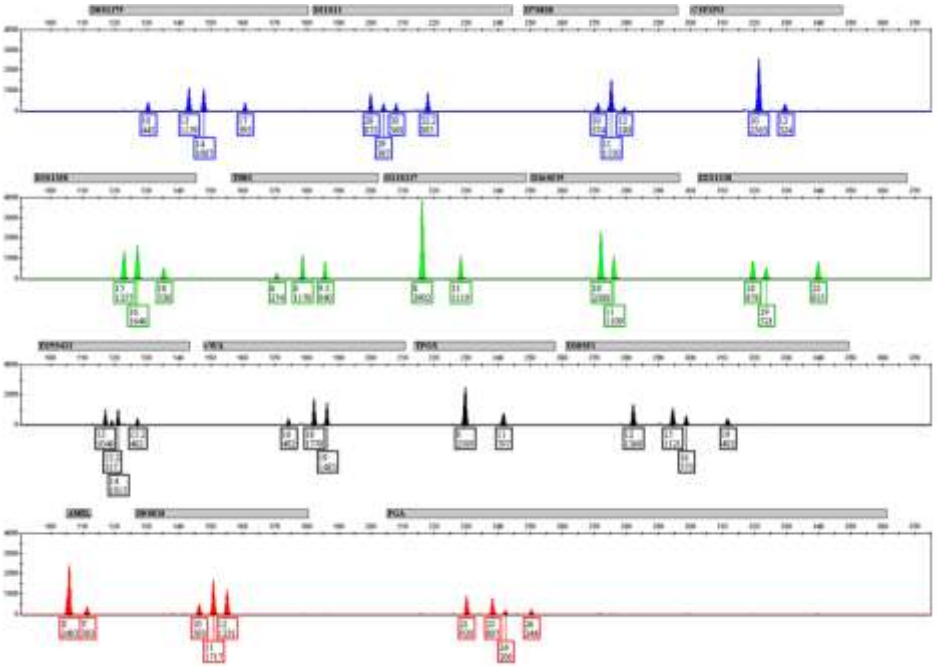
Gender identification + 25 autosomal STR loci in a single amplification

Component D

3 part A

1 part C



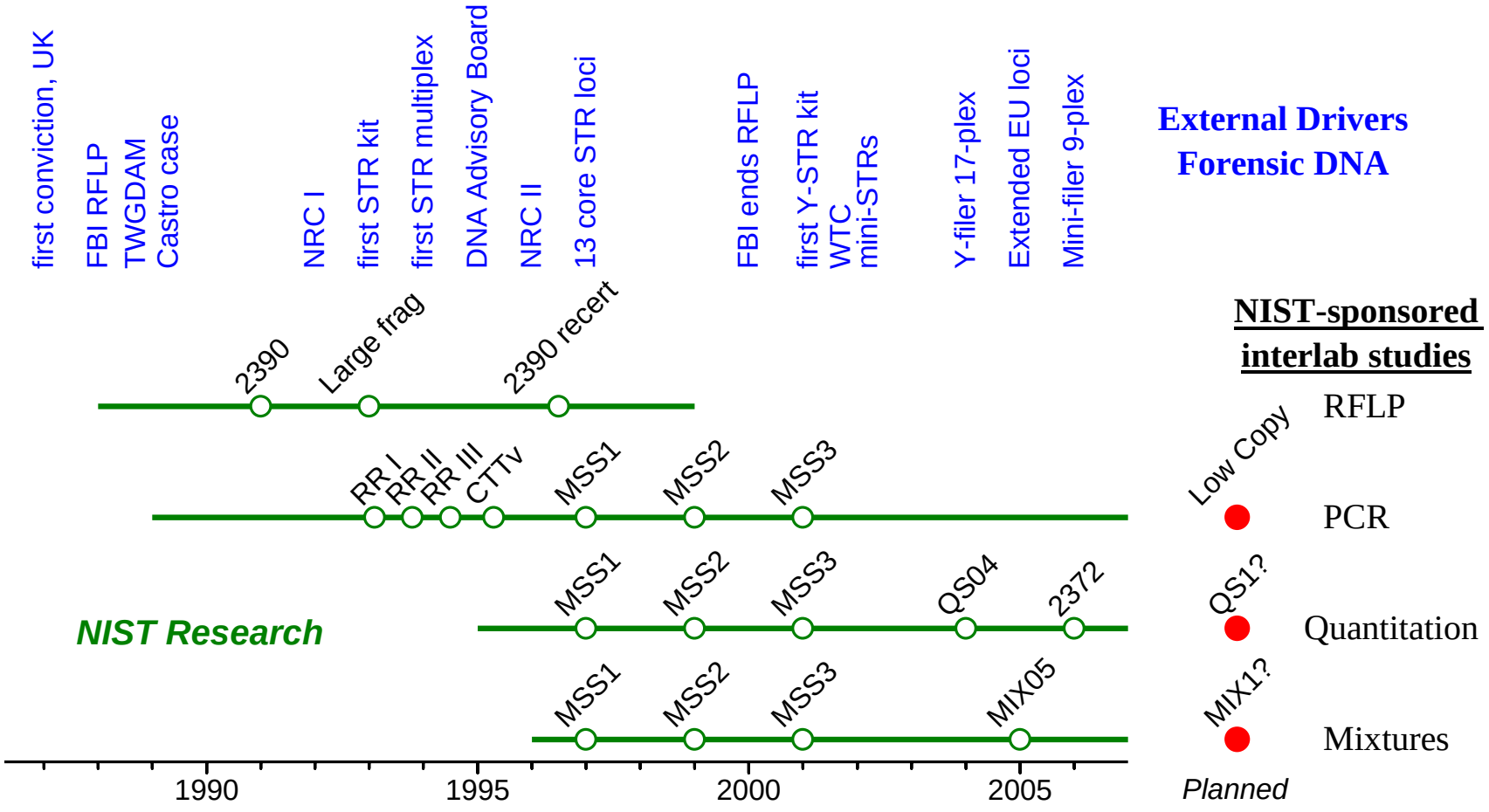


The certified ratio for Component D, the mass of Component A relative to that of Component C, is **3.1 ± 0.1 Component A / Component C.**

NIST-Sponsored Interlab Studies


Margaret Kline


Dave Duewer



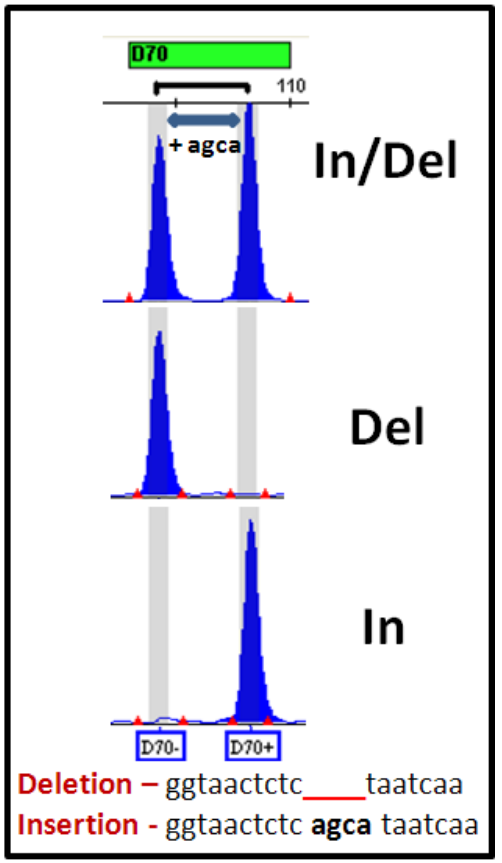
13 interlaboratory studies conducted over the past 20 years

Insertion/Deletion (InDel) Markers

Manuel **Fonde**vila Alvarez
Guest Researcher from Spain

Main Points:

- InDels (insertion-deletion) or DIPs (deletion-insertion polymorphisms) are short length polymorphisms, consisting of the presence or absence of a short (typically 1-50 bp) sequence
- Like SNPs, InDels have low mutation rate (value to kinship analysis), small amplicon target sizes (value with degraded DNA), and can be highly multiplexed
- Can be analyzed on CE instruments like STRs
- Studied commercial 30plex (Qiagen DIPlex) and a home-brew 38plex in U.S. population samples



Presentations/Publications:

- FSI Genetics Suppl. Series 2011 article
- ISFG 2011 poster and ISHI 2011 presentation

STR Kit Concordance Testing

Becky Hill

Main Points:

- When different primer sets are utilized, there is a concern that allele dropout may occur due to primer binding site mutations that impact one set of primers but not another
- To test SRM 2391b/2391c (PCR-based DNA Profiling Standard) components with all new STR multiplex kits and verify results against certified reference values
- To gain a better understanding of primer binding site mutations that cause null alleles

If no primer binding site mutations

Set 1 Amplicons Set 2 Amplicons

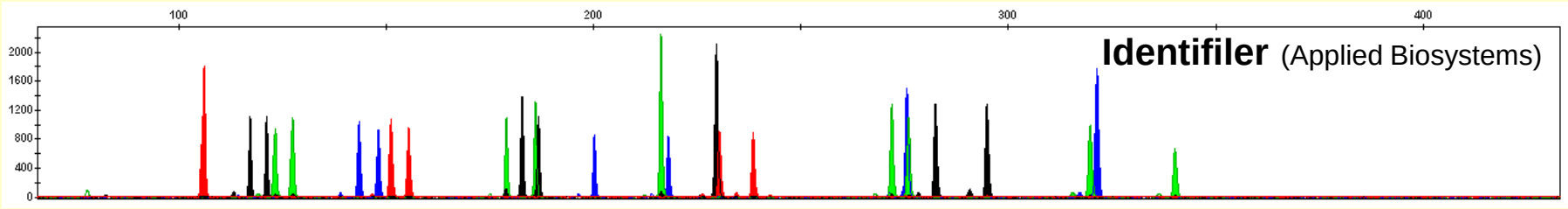
If a primer binding site mutation exists

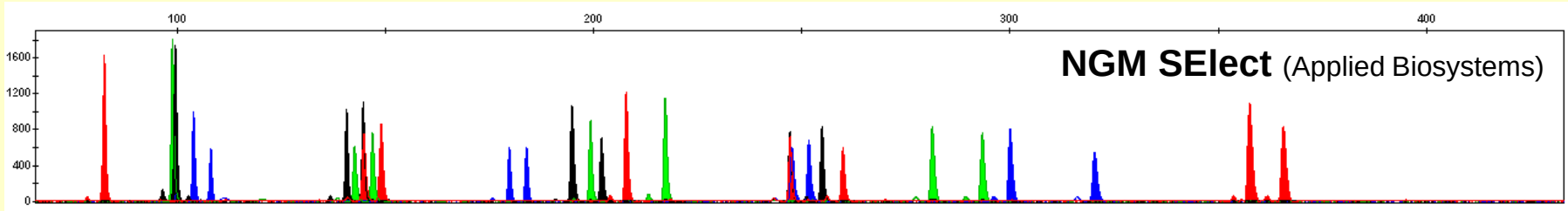
Set 1 Amplicons Set 2 Amplicons

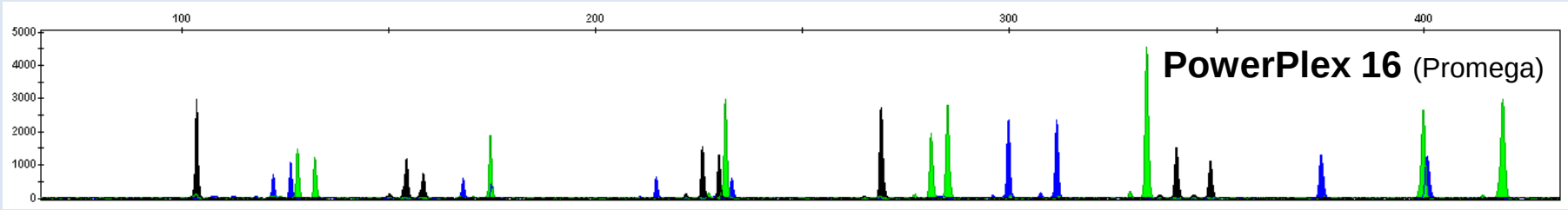
Presentations/Publications:

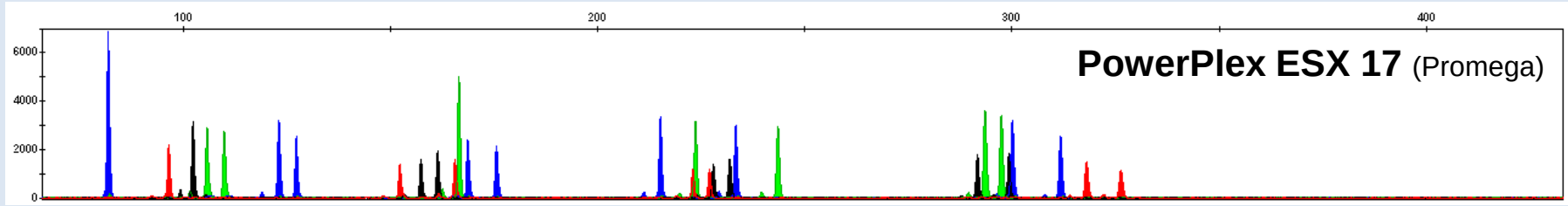
- *Profiles in DNA* article (Hill et al. 2010)
- ISFG 2011 and ISHI 2011 posters (Hill et al.)

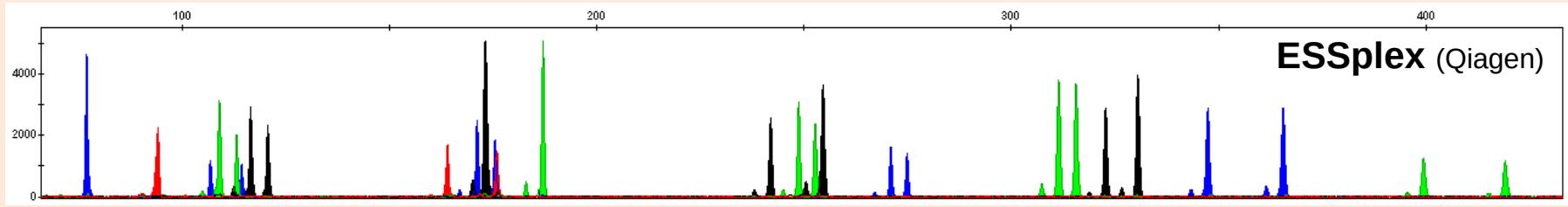
Same DNA Sample Tested with Five STR Kits











Kit Concordance Comparisons

Kits compared	Samples	Loci compared	Comparisons	# Differences	Concordance (%)
SGM-ID	1436	11	15,796	1	99.994
ID-ProPlus	1427	10	14,270	1	99.993
ID-IDplex	669	16	10,704	19	99.822
ID-PP16	662	14	9,268	4	99.957
ID-MiniFiler	1308	9	11,772	27	99.771
SGM-NGM	1436	11	15,796	4	99.975
ID-NGM	1449	11	15,939	3	99.981
ProPlus-NGM	1427				
SGM-ESI	1436				
ProPlus-ESX	1427				
ESI-ESX	1455				
ESI-ESSplex	1445				
ESX-ESSplex	1445				
ESI-NGMSElect	715				
ESX-NGMSElect	715				
ESS-NGMSElect	663	17	11,271	17	99.849
		TOTAL	240,156	186	99.923

> 1 million allele comparisons

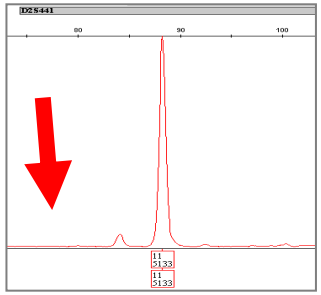
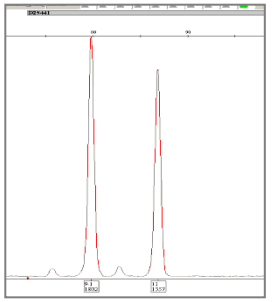
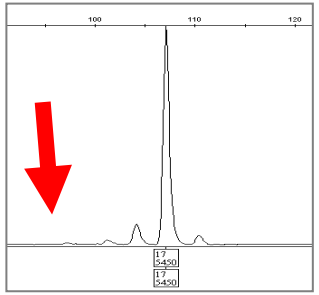
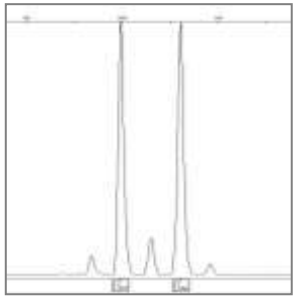
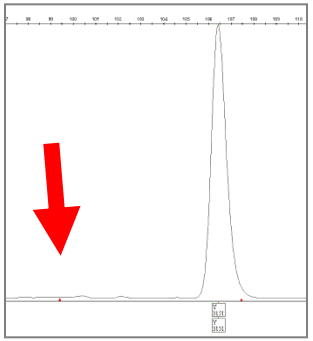
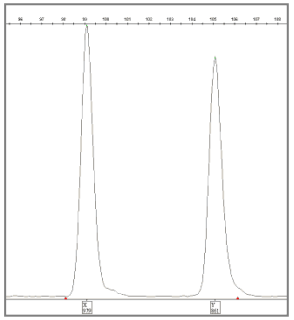
>1100 differences observed

~99.9% concordance

(many corrected now)

*Kits (except Identifiler) were kindly provided by **Applied Biosystems, Promega, and Qiagen** for concordance testing performed at NIST*

Extra (Degenerate) Primers Added with NGM SElect

	NGM (original)		NGM SElect and NGM’	
D2S441 9.1 allele missing in 7 Asians		11,11		9.1,11
D22S1045 15 allele missing in 4 samples		17,17		15,17
Amelogenin X allele missing in 3 samples		Y,Y		X,Y

Variant STR Allele Sequencing

Margaret Kline

Main Points:

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

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

STR sequence analysis for characterizing normal, variant, and null alleles

Margaret C. Kline *, Carolyn R. Hill, Amy E. Decker [1], John M. Butler

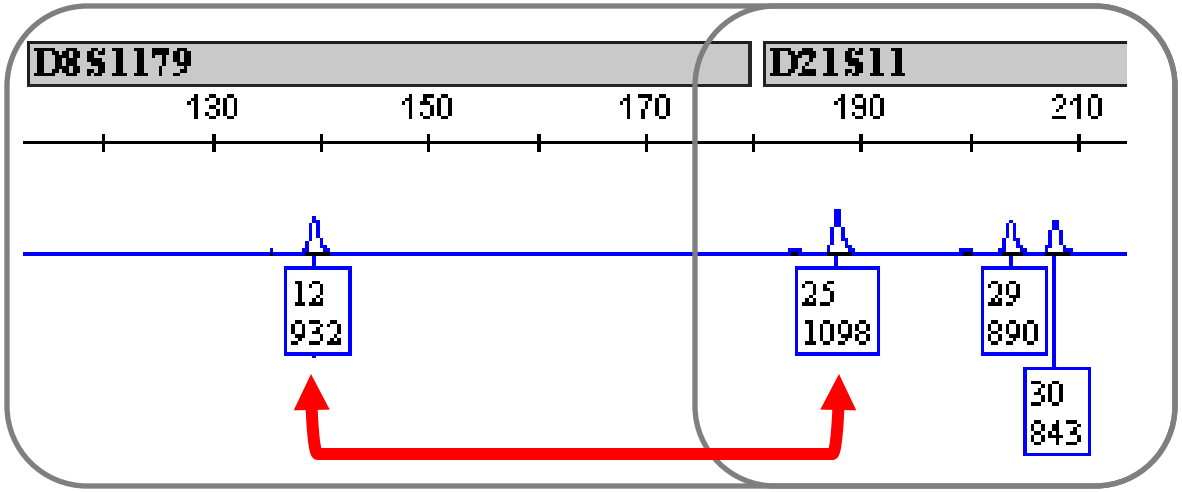
National Institute of Standards and Technology, 100 Bureau Drive, M/S 8312, Gaithersburg, MD 20899, USA

Presentations/Publications:

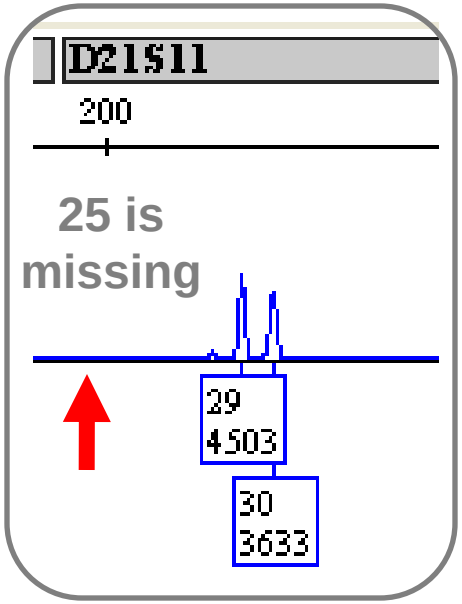
- FSI Genetics article (Aug 2011) and numerous talks

How Do You Characterize Your Tri-Allelic Patterns?

Identifiler



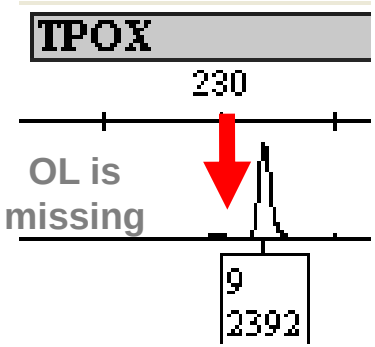
PowerPlex 16 HS

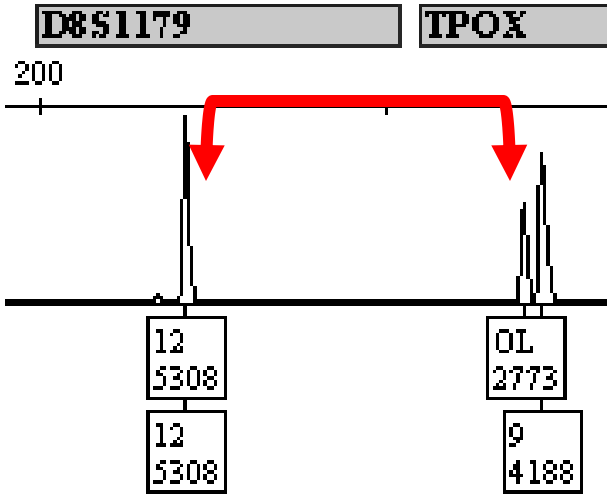


You re-amplify it…
It's Reproducible!

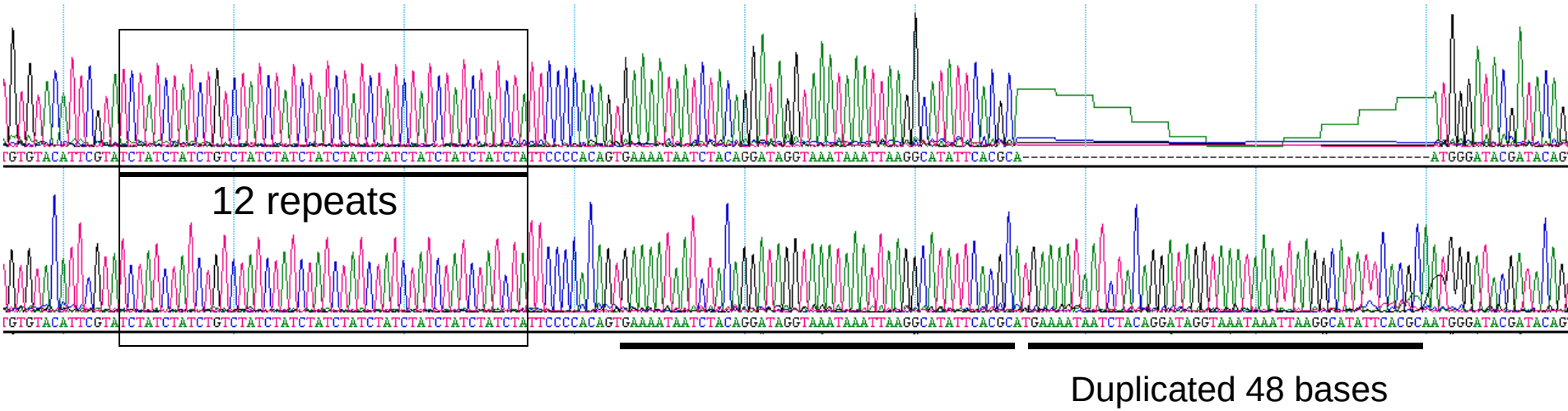
Check STRBase…
It has never been observed before!

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





D8S1179 12,“24”



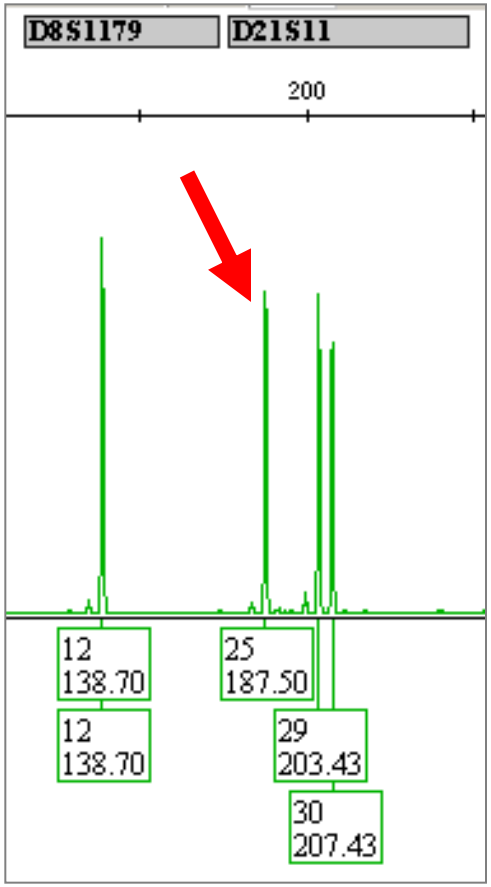
Allele 12 : $[TCTA]_2$ TCTG $[TCTA]_9$

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

Result with This Large D8S1179 Allele

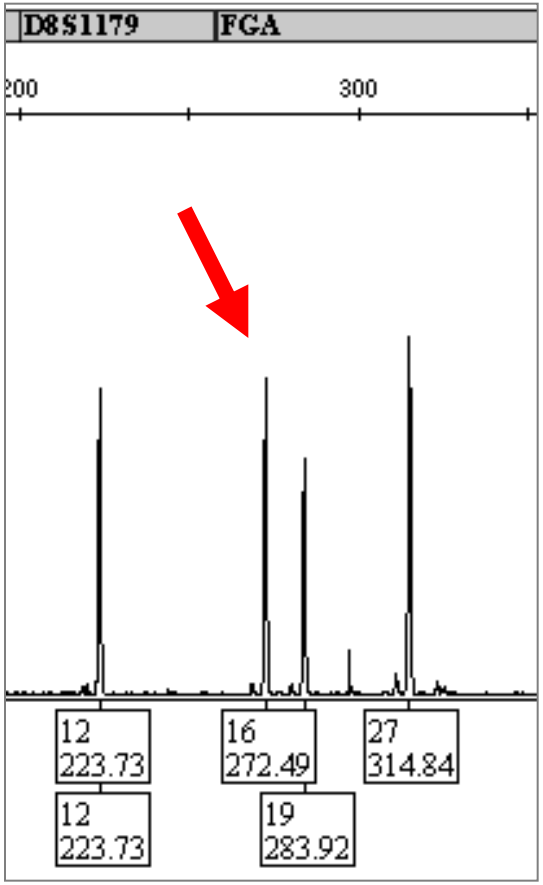
Using European STR Kits

NGM SElect



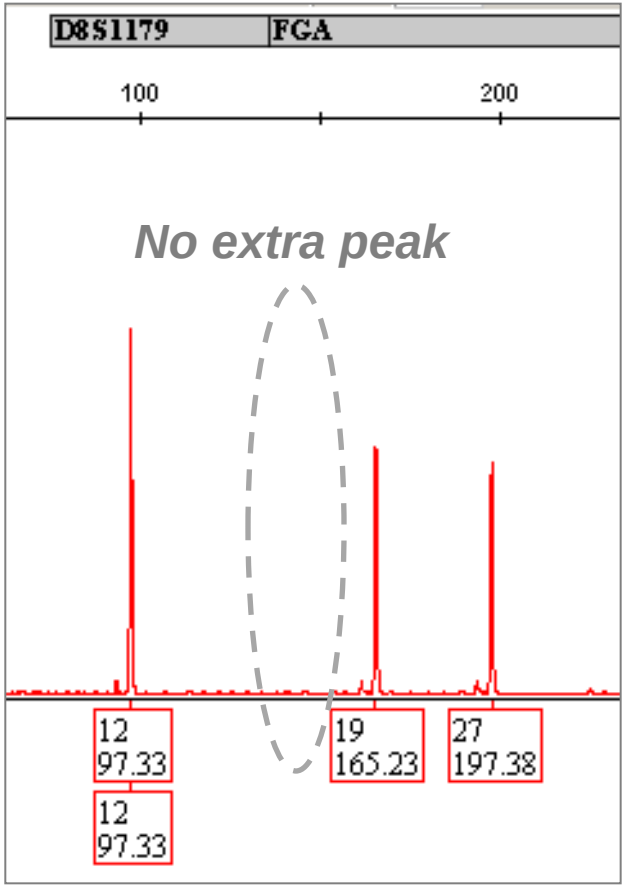
False D21S11 tri-allele

PP ESX 17



False FGA tri-allele

PP ESI 17



Reverse primer internal to duplicated flanking region

Recent Training Workshops


John Butler


Mike Coble

- AAFS (February 22, 2011)
  - **Mixture Interpretation (with 6 other speakers)**
- ISFG (August 30, 2011)
  - **CE Fundamentals and Troubleshooting**
- Int. Symp. Human Ident. (October 3, 2011)
  - **Mixture Interpretation (with Boston University)**
- Int. Symp. Human Ident. (October 6, 2011)
  - **Troubleshooting Laboratory Systems**

Slide handouts available at
http://www.cstl.nist.gov/strbase/training.htm

Mixture Workshop (Promega ISHI 2010)

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

Regional workshops presented in FL, TX, MI, and AZ (April – June 2011)
Updated mixture workshop presented at ISHI 2011 (October 3, 2011)

Handout >200 pages

Literature list of >100 articles

13 Modules Pres

Introductio
SWGD (John)
olds (Catherine)
astic effects (Robin)
eak height ratios (Charlotte)
Number of contributors (John)
Mixture ratios (John)
Mixture principles (Charlotte)
Statistics (Mike)

Case Example 1 (Robin)
Case Example 2 (Charlotte)
Case Example 3 (John)

Catherine Grgicak Boston U.
Mike Coble NIST
Robin Cotton Boston U.
John Butler NIST
Charlotte Word Consultant

NIJ Grant to Boston University
funded ~150 state & local lab analysts to attend

TrueAllele Mixture Software Evaluation

Mike Coble

Main Points:

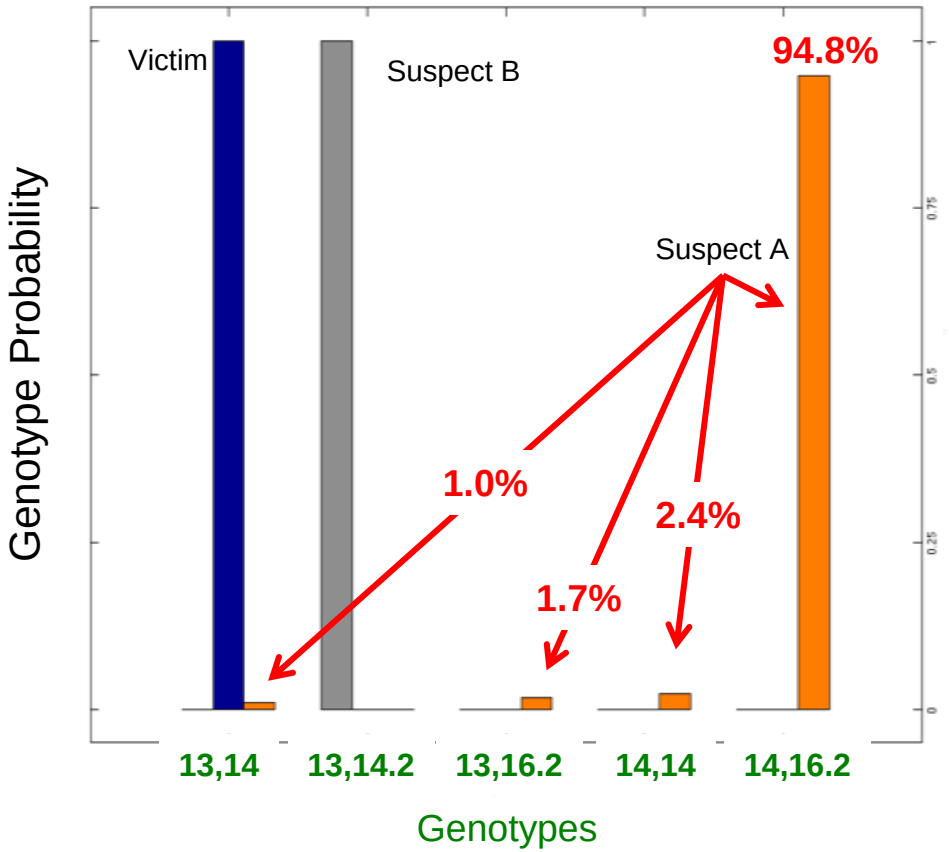
- Exploring the capabilities and limitations of a probabilistic genotyping approach
- Studying TrueAllele software with a number of different types of mixtures (including low-level and 3-4 person mixtures)
- Work being performed at NIST independently of Cybergenetics

Presentations/Publications:

- ISFG 2011 presentation
- ISHI 2011 mixture workshop

D19S433 result from one replicate of 50,000 simulations

3 person mixture conditioning on the victim



See also Perlin et al. (2011) Validating TrueAllele DNA mixture interpretation. *J. Forensic Sci.* 56(6):1430-1447

Rapid PCR and Rapid DNA Testing


Pete Vallone

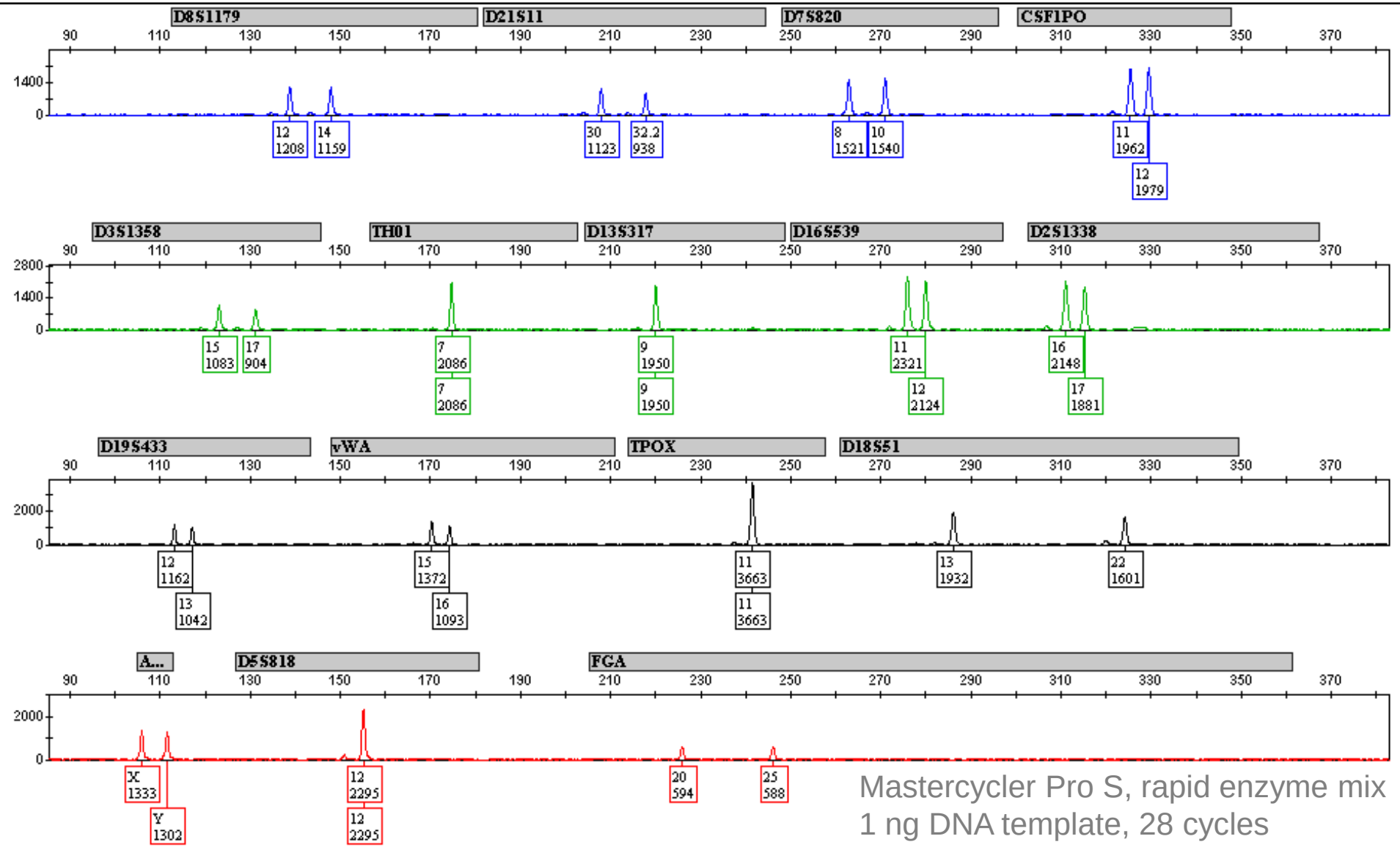
Main Points:

- **Performing research on reducing the total time required** for STR typing
  - Focusing on the multiplex amplification of commercial STR kits with faster polymerases and thermal cyclers
  - Single-source reference samples (sensitivity > 200 pg)
- **Designing testing plans for rapid DNA typing devices**
  - NIST will be examining rapid DNA instruments with FBI collaboration
- **Exploring direct PCR protocols** with FTA and 903 papers

Presentations/Publications:

- Vallone et al. (2008) FSI Genetics - on rapid PCR
- ISFG 2011 and ISHI 2011 presentations by Tom Callaghan (FBI)
- ISFG 2011 presentation and poster on direct PCR

Full Identifiler STR Profile with **19 min PCR**



Mastercycler Pro S, rapid enzyme mix
1 ng DNA template, 28 cycles

ABI 3500 Validation Studies

Erica Butts

Main Points:

- The 3500 has proven to be reliable, reproducible and robust in our hands – we have provided feedback to ABI to improve use
- Produces excellent DNA sequencing results
- Signal strength is different compared to ABI 3130xl and requires studies to set analytical and stochastic thresholds
- **Dye-specific analytical thresholds** resulted in less allelic and full locus dropout than applying one analytical threshold to all dyes
- RFID tracking decreases flexibility in our research experience

Presentations/Publications:

- MAAFS talk (May 2011)
- ABI road show talks (July & Aug 2011)
- ISFG presentation (Sept 2011)
- ISHI poster (Oct 2011)

Future Projects Planned

- New book in progress on interpretation issues
- Additional mixture software evaluation
- Rapidly mutating Y-STR loci (European collaboration)
- More concordance testing with new STR kits
- PLEX-ID mass spec validation with mtDNA base composition (FBI collaboration)
- Rapid DNA test device evaluation (FBI collaboration)
- Exploration of Next-Generation Sequencing
- Digital PCR for human DNA quantitation

ABI 3500 Genetic Analyzer

Status Update on March 2011

Open Letter to Applied Biosystems

Open Letter to Applied Biosystems on Concerns with ABI 3500

- **3/14/11 - emailed ~900 forensic DNA scientists** (SWGDAM, forens-dna, ENFSI, EDNAP) inviting them to sign onto a letter that will be sent to Applied Biosystems expressing concern with ABI 3500
- **Very positive response with 101 who agreed to sign the letter**
- Letter was sent March 31 to the president of ABI and scientists involved with the ABI 3500
- **Community will be notified of ABI's response**

Concerns Expressed in Open Letter

- RFID tags
- New .hid file structure requires new software
- Short shelf life of reagents – would like to see data for expiration times

Hopefully a change will result…

A desire for greater communication with the community – the 3500 FAQ sheet is a good start but does not directly address all of the concerns raised

Brief Timeline of Events

- NIJ requested NIST to explore capabilities, limitations, and cost of ABI 3500 instrument and reagents (May 2010)
- NIST presentations to NIJ (Dec 2010) and SWGDAM (Jan 2011)
- **Open letter support solicited and sent to ABI** (Mar 2011)
- Further discussions between NIST and ABI (Apr-Sept 2011)
- At the Promega ISHI meeting (Oct 2011), ABI announced through a poster at their booth that **polymer and buffer expiration dates will no longer be a hard stop** but only a warning with the future Windows 7 software upgrade

Since May 2011, Erica Butts has presented several validation presentations on our ABI 3500 work – these are available on STRBase

Characterizing New STR Loci


John Butler


Becky Hill

Main Points:

- In April 2011, the FBI announced plans to expand the core loci for the U.S. beyond the current 13 CODIS STRs

- Our group is collecting U.S. population data on new loci and characterizing them to aid understanding of various marker combinations

- We are collecting all available information from the literature on the 24 commonly used autosomal STR loci

Presentations/Publications:

- AAFS 2011 presentation
- Hill et al (2011) *FSI Genetics* (Aug 2011 issue)
- Butler & Hill (2011) *Forensic Sci Rev* (submitted)
- Hares (2011) Expanding the U.S. core loci… *FSI Genetics* (in press)

April 2011 *FSI Genetics* Letter to the Editor

Describing Intention to Expand the CODIS Core Loci in the United States

CODIS Core Loci Working Group
Formed in May 2010 to make recommendations to FBI CODIS Unit

Douglas Hares (Chair) – FBI
John Butler – NIST
Cecelia Crouse – FL PBSO
Brad Jenkins – VA DFS
Ken Konzak – CA DOJ
Taylor Scott – IL SP

ARTICLE IN PRESS

Forensic Science International: Genetics xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Letter to the Editor

Expanding the CODIS core loci in the United States

Dear Editor:

After over a decade of operation, the National DNA Index System (NDIS) continues to grow in importance and size [1]. While the STR DNA technology has remained relatively consistent, other key aspects of the NDIS program have been reevaluated and revisions implemented. For example, based upon recommendations of the Scientific Working Group on DNA Analysis Methods, the Director of the Federal Bureau of Investigation (FBI) issued revised Quality Assurance Standards (QAS) for Forensic DNA

major reasons for expanding the CODIS core loci in the United States:

(1) To reduce the likelihood of adventitious matches [7] as the number of profiles stored at NDIS continues to increase each year (expected to total over 10 million profiles by the time of this publication). There are no signs that this trend will slow down as States expand the coverage of their DNA database programs and increase laboratory efficiency and capacity.

(2) To increase international compatibility to assist law enforcement data sharing efforts.

(3) To increase discrimination power to aid missing persons cases.

Points for Consideration with Expanded CODIS Core Loci

- Why expand the number of CODIS core loci?
- What can we learn from European Standard Set (ESS) expansion experience?
- What are the proposed U.S. expanded loci?
- What concerns have been raised in recent public criticism?
- **What data exist behind decisions made so far** and what additional data are there for consideration to help address concerns raised?
- Summary

Three major reasons for expanding the CODIS core loci in the United States

D.R. Hares (2011) *Forensic Sci. Int. Genet.*

- **To reduce the likelihood of adventitious matches** as the number of profiles stored at NDIS continues to increase each year

- **To increase international compatibility** to assist law enforcement data sharing efforts

- **To increase discrimination power to aid missing persons cases**

Adventitious Matches

- The only published account of a false match from a DNA database came in 1999 when the UK database then consisting of 660,000 profiles with only 6 STR loci (SGM assay) lead to a "hit" between two individuals whose 6-locus random match probability was 1 in 37 million (R. Willing, *USA Today*, Feb 8, 2000; "Mismatch calls DNA test into question").

- Further testing with four additional STRs (SGM Plus loci) showed that the samples were from different individuals. **The UK expanded the number of core loci from 6 to 10 with the adoption of the SGM Plus kit to try and prevent another adventitious match**.

- The growth of DNA databases necessitates the inclusion of additional loci to avoid this problem.

For further information, see D.N.A. Box 8.3 in Butler, J.M. (2012) *Advanced Topics in Forensic DNA Typing: Methodology*, p. 251

International Comparability

Currently there are 24 autosomal STR markers present in commercial kits

ESS = European Standard Set

13 CODIS loci

U.S.	Europe
TPOX	
CSF1PO	
D5S818	
D7S820	
D13S317	
FGA	FGA
vWA	vWA
D3S1358	D3S1358
D8S1179	D8S1179
D18S51	D18S51
D21S11	D21S11
TH01	TH01
D16S539	D16S539
D2S1338	D2S1338
D19S433	D19S433
Penta D	
Penta E	

7 ESS loci

5 loci adopted in 2009 to expand to 12 ESS loci

D12S391
D1S1656
D2S441
D10S1248
D22S1045
SE33 → Core locus for Germany

3 miniSTR loci developed at NIST

Locus used in China ← **D6S1043**

What can we learn from European Standard Set (ESS) expansion experience?

PROFILES IN DNA

March 2009 issue

THE ESS LOCI

Expansion of the European Standard Set of DNA Database Loci—the Current Situation

Peter M. Schneider
Institute of Legal Medicine, University Hospital of Cologne, Germany

http://www.promega.com/resources/articles/profiles-in-dna/2009/expansion-of-the-european-standard-set/

ELSEVIER

Forensic Science International 119 (2001) 225–231

Forensic Science International

www.elsevier.com/locate/forsciint

A brief history of the formation of DNA databases in forensic science within Europe

Peter D. Martin[a,*], Hermann Schmitter[b], Peter M. Schneider[c]

[a]32 Oakfield Gardens, Kent, BR3 3AZ Beckenham, UK
[b]Serology Department, Bundeskriminalamt, Thaerstr. 11, 65193 Wiesbaden, Germany
[c]Institute of Legal Medicine, Johannes Gutenberg University, Am Pulverturm 3, 55131 Mainz, Germany

Received 10 September 2000; accepted 14 November 2000

Interpol European Working Party on DNA Profiling

Schmitter[3], Uwe Schleenbecker[3], Janos Woller[4], Paola Montagna[5],
en Wolf[7], Nils Moe[8], Jan Matusek[9], José Andradas Heranz[10],
Serra Utrilla[10], William Grahamslaw[11], Peter Fifka[11], Mark Branchflower[12]
uxelles, Belgium

Forensic Science International 119 (2001) 232–238

Forensic Science International

www.elsevier.com/locate/forsciint

Criminal DNA databases: the European situation

Peter M. Schneider[a,*], Peter D. Martin[b]

[a]Institute of Legal Medicine, Johannes Gutenberg University, Am Pulverturm 3, 55131 Mainz, Germany
[b]Former Deputy Director, Metropolitan Police Forensic Science Laboratory, London, UK

Received 10 September 2000; accepted 14 November 2000

DNA Databases for Offender Id
The Need for Technical, Legal an

Peter M. Schneider
Institute of Legal Medicine, Johannes Gutenberg University,
Am Pulverturm 3, 55131 Mainz, Germany

http://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/2nd-eshi-oral-presentations/

European Expansion Efforts

More loci added as databases grew…

European Standard Set (ESS) of Loci
1998: 4 STRs
1999: 7 STRs
2009: 12 STRs
(but 15 with most kits)

1995
UK National Database launched

UK began with 6 STRs (SGM)

1998
Leriche et al. (1998)
Initial Interpol ESS selection (4 STRs)

1999
ESS increased to 7 STRs; UK went to 10 STRs (SGM Plus)

2001
Martin et al. (2001)
Multiple countries have established national databases

2004
Degraded DNA interlab study (EDNAP)

2005
Prüm treaty
(data sharing)

EDNAP/ENFSI recommend new loci

2006
Gill et al. (2006a, 2006b)
Letters to editor (1) & (2) announcing proposed new loci - miniSTRs advocated

Dixon et al. (2006)
EDNAP interlab published

2007
Prototype kits developed & tested

2009
ESS expanded to 12 loci

2010
NGM and NGM SElect; Qiagen kits

PowerPlex ESI/ESX 16/17 kits released

2011
Implementation required

With expanded loci selections, focus is on casework capabilities not familial searching potential

Lessons from European ESS Expansion

- **Data studies should drive decisions**
  - Interlaboratory study with degraded DNA (Dixon et al. 2006 article was key)
- Casework capabilities are a primary goal
  - miniSTRs and desire for kits with ability to overcome inhibitors
- **Initial locus selection announced through Letters to the Editor** of the leading forensic DNA journal (Gill et al. 2006a, 2006b)
- Companies responded with prototype kits for evaluation
- Expanded ESS loci were selected and voted upon after data review by ENFSI labs **(4 years after initial recommendations were made)**
- EU adopted recommendations of ENFSI
- Commercial kits became available to meet expanded ESS requirements
- Population data gathered and software developed
- European labs must be compliant by Nov 30, 2011 (**2 years after adoption**)
- Casework capabilities not familial searching potential were the intent of the core loci selection

EDNAP Study Showed Value of miniSTRs

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Forensic Science International 164 (2006) 33–44

Forensic Science International

www.elsevier.com/locate/forsciint

Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (EDNAP) exercise

L.A. Dixon [a,*], A.E. Dobbins [a], H.K. Pulker [a], J.M. Butler [b], P.M. Vallone [b], M.D. Coble [b], W. Parson [c], B. Berger [c], P. Grubwieser [c], H.S. Mogensen [d], N. Morling [d], K. Nielsen [d], J.J. Sanchez [d], E. Petkovski [e], A. Carracedo [f], P. Sanchez-Diz [f], E. Ramos-Luis [f], M. Briõn [f], J.A. Irwin [g], R.S. Just [g], O. Loreille [g], T.J. Parsons [g], D. Syndercombe-Court [h], H. Schmitter [i], B. Stradmann-Bellinghausen [j], K. Bender [j], P. Gill [a]

[a] *The Forensic Science Service, Research and Development, Trident Court, Birmingham, UK*
[b] *National Institute of Standards and Technology, Gaithersburg, MD, USA*
[c] *Institute of Legal Medicine, Innsbruck Medical University, Austria*

“Recently, there has been much debate about what kinds of genetic markers should be implemented as new core loci that constitute national DNA databases. The choices lie between conventional STRs, ranging in size from 100 to 450 bp; mini-STRs, with amplicon sizes less than 200 bp; and single nucleotide polymorphisms (SNPs)…Results were collated and analysed and, in general, **mini-STR systems were shown to be the most effective**…”

Data Driven Decisions

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Forensic Science International 156 (2006) 242–244

Forensic Science International

www.elsevier.com/locate/forsciint

Short communication

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

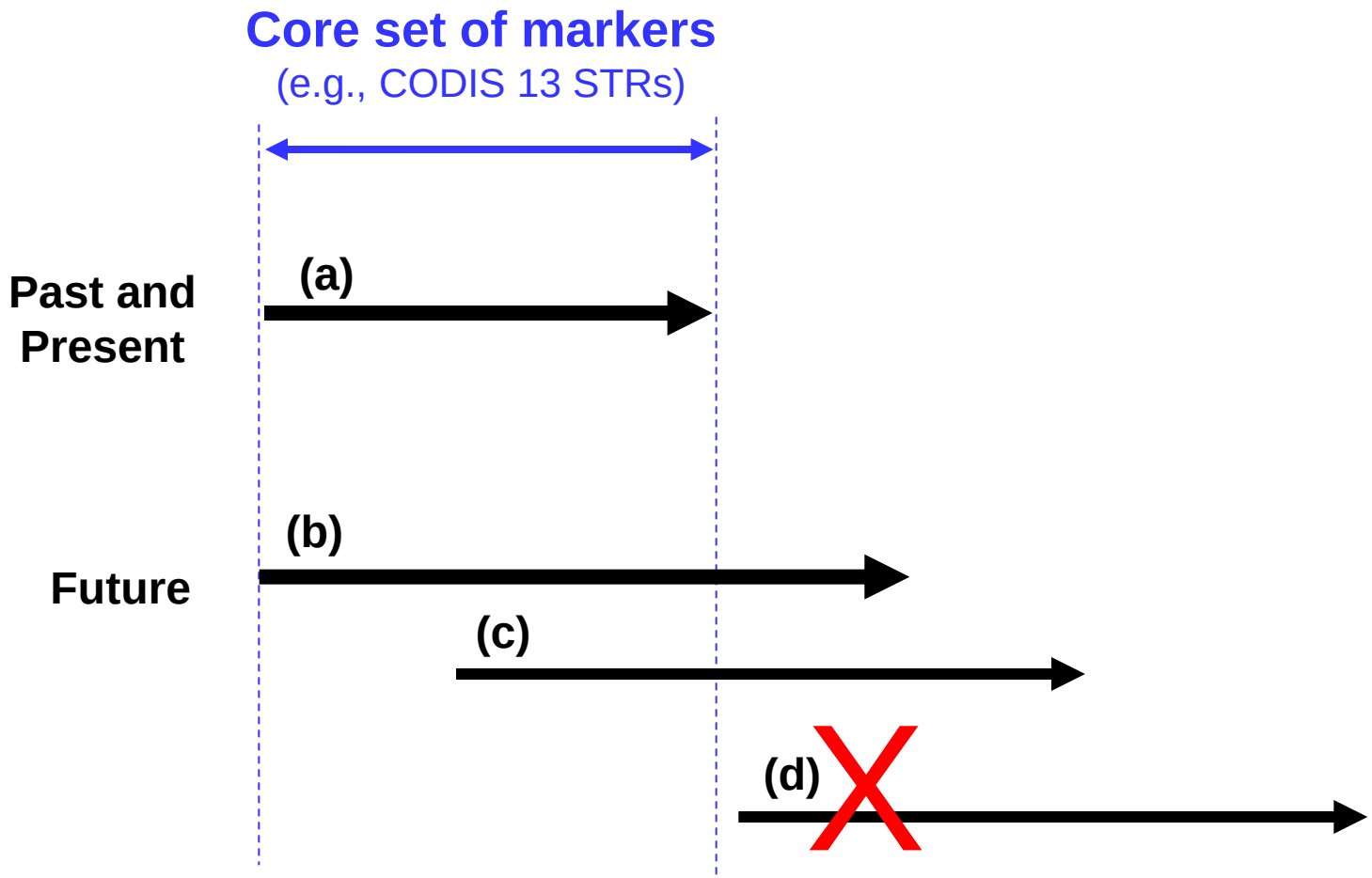
Peter Gill [a,*], Lyn Fereday [b], Niels Morling [c], Peter M. Schneider [d]

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

Received 25 May 2005; accepted 26 May 2005
Available online 5 July 2005

"Following a recent meeting by the ENFSI and EDNAP groups on the 4–5 April, 2005, in Glasgow, UK, it was unanimously agreed that the process of standardization within Europe **should take account of recent work that unequivocally demonstrated that chance of obtaining a result from a degraded sample was increased when small amplicons (mini-STRs) were analysed**…"

Possible scenarios for extending sets of genetic markers to be used in national DNA databases



Maintaining connection to legacy data is essential

Figure 8.4 in Butler (2012) *Advanced Topics in Forensic DNA Typing: Methodology*, p. 243

Proposed Expanded CODIS Core Loci

D.R. Hares (2011) *Forensic Sci. Int. Genet.*

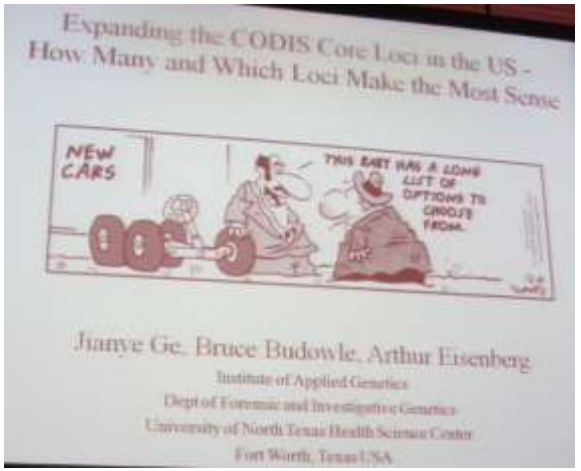
21 required loci
Amelogenin (for sex-typing)
19 autosomal STRs
1 Y-STR (DYS391)

	Locus
Section A (required)	Amelogenin
	D18S51
	FGA
	D21S11
	D8S1179
	vWA
	D13S317
	D16S539
	D7S820
	TH01
	D3S1358
	D5S818
	CSF1PO
	D2S1338
	D19S433
	D1S1656
	D12S391
	D2S441
	D10S1248
	Penta E
	DYS391

	Locus
Section B (in order of preference)	TPOX
	D22S1045
	SE33
	Penta D

Current CODIS 13 loci in red font

Recent Public Criticism of Efforts to Expand the CODIS Core Loci in the U.S.



October 4, 2011 Presentation at Promega's International Symposium on Human Identification

BBC Mobile NEWS

October 16, 2011 Follow-up article by BBC News on ISHI presentation

http://ishinews.com/

FBI's DNA Database Upgrade Plans Come Under Fire

OCTOBER 18, 2011 9:00 AM FEATURED ARTICLE, GENERAL SESSION UPDATES, RECENT ARTICLE

By: Paul Rincon, Science editor Published October 16, 2011 by the BBC News Website A major upgrade of the Federal Bureau of Investigation's (FBI) DNA database system has come under fire from members of the forensic science community. The Codis system is used to generate the genetic profiles stored in the US national DNA [...]

http://www.bbc.co.uk/news/science-environment-15311718

Concerns Raised in Public Criticisms of Expanded CODIS Core Loci Selection

- **Not enough data behind decisions** – need more community involvement rather than a small committee making decisions
- **Casework needs** should drive decisions
- **Large loci fail in casework samples and should be avoided** – miniSTR capabilities are preferred
- **Large multiplexes** may adversely impact performance
- **DYS391 is a poor choice** and AMEL Y nulls are not a significant concern
- **Y-STRs should be included** as core loci to benefit familial searches of the future
- **No definition of performance goals** are provided

Some Points for Consideration

- There is only so much room (“electrophoretic real-estate”) available with 5-dyes and amplicons <500 bp
  - A sixth-dye may help expand multiplex capabilities but ABI 3500 instruments are not yet available in all labs
- If loci have smaller allele ranges, then more loci can fit into this limited space
  - For example, FGA could be replaced by 2-3 smaller loci
  - Smaller allele ranges typically correlate with lower mutation rates
- Improved chemistry in kits makes megaplexes feasible
- Extensive NIST concordance studies are aiding companies to avoid primer binding site mutations and produce better kits
- Y-STR loci are only useful for male DNA samples

What data exist behind decisions made so far and what additional data are there for consideration to help address concerns raised?

The 11 STR Loci Beyond the CODIS 13

STR Locus	Location	Repeat Motif	Allele Range*	# Alleles*
D2S1338	2q35	TGCC/TTCC	10 to 31	40
D19S433	19q12	AAGG/TAGG	5.2 to 20	36
Penta D	21q22.3	AAAGA	1.1 to 19	50
Penta E	15q26.2	AAAGA	5 to 32	53
D1S1656	1q42	TAGA	8 to 20.3	25
D12S391	12p13.2	AGAT/AGAC	13 to 27.2	52
D2S441	2p14	TCTA/TCAA	8 to 17	22
D10S1248	10q26.3	GGAA	7 to 19	13
D22S1045	22q12.3	ATT	7 to 20	14
SE33	6q14	AAAG‡	3 to 49	**178**
D6S1043	6q15	AGAT/AGAC	8 to 25	25

5 new European loci

**Allele range and number of observed alleles from Appendix 1, J.M. Butler (2011) Advanced Topics in Forensic DNA Typing: Methodology; ‡SE33 alleles have complex repeat structure*

Loci sorted on Probability of Identity (P_I) values

STR Locus	Alleles Observed	Genotypes Observed	Het. (obs)	P_I value N = 938
SE33	53	292	0.9360	0.0069
Penta E*	20	114	0.8799	0.0177
D2S1338	13	68	0.8785	0.0219
D1S1656	15	92	0.8934	0.0220
D18S51	21	91	0.8689	0.0256
D12S391	23	110	0.8795	0.0257
FGA	26	93	0.8742	0.0299
D6S1043*	25	91	0.8627	0.0343
Penta D*	16	71	0.8754	0.0356
D21S11	25	81	0.8358	0.0410
D19S433	16	76	0.8124	0.0561
D8S1179	11	45	0.7878	0.0582
vWA	11	38	0.8060	0.0622
D7S820	11	32	0.8070	0.0734
TH01	8	24	0.7580	0.0784
D16S539	9	28	0.7825	0.0784
D13S317	8	29	0.7655	0.0812
D10S1248	12	39	0.7825	0.0837
D2S441	14	41	0.7772	0.0855
D3S1358	11	30	0.7569	0.0873
D22S1045	11	42	0.7697	0.0933
CSF1PO	9	30	0.7537	0.1071
D5S818	9	34	0.7164	0.1192
TPOX	9	28	0.6983	0.1283

24 STR Loci in STR kits rank ordered by their variability

Better for mixtures (more alleles seen)

There are several loci more polymorphic than the current CODIS 13 STRs

Better for kinship (low mutation rate)

Concern: Large loci fail in casework samples and should be avoided – miniSTR capabilities are preferred

- We agree that miniSTRs (smaller amplicons) work best with degraded DNA that is often present in casework samples
- **How often are high molecular weight loci failing?**
- What data exist on success rates of loci for profiles stored in Forensic Index of CODIS based on PCR product size?

Palm Beach Sheriff's Office Crime Lab

LDIS Forensic Unknowns – PowerPlex 16 data

Single-source

2,452 profiles total

- Loss of Penta D: 633
- Loss of Penta E: 323
- Loss of FGA: 202
- Loss of all 3 loci: 130

130/2452 = 5.3%

FGA loss = 8.2%

Mixtures

841 profiles total

- Loss of Penta D: 297
- Loss of Penta E: 296
- Loss of FGA: 179
- Loss of all 3 loci: 55

55/841 = 6.5%

FGA loss = 21.3%

Larger loci are lost in a fraction of casework samples…

Data courtesy of Cecelia Crouse & Tara Sessa (PBSO)

Additional Data from VA and CA

	Virginia	California
# Forensic Unknowns (single-source profiles)	13,488	37,024
No FGA (largest of current CODIS 13 core loci)	68 (**0.5%**)	1,936 (**4.5%**)
Profiles missing at least one locus	1,609 (12%)	4,440 (12%)

Data courtesy of George Li, Brad Jenkins, and Ken Konzak

Will Performance with Large Multiplexes Be Adversely Impacted with Additional Loci Added?

- **There has been significant improvement in kit development in recent years**
  - In addition, 6-dye capability of ABI 3500 instruments may play a role in future kits…

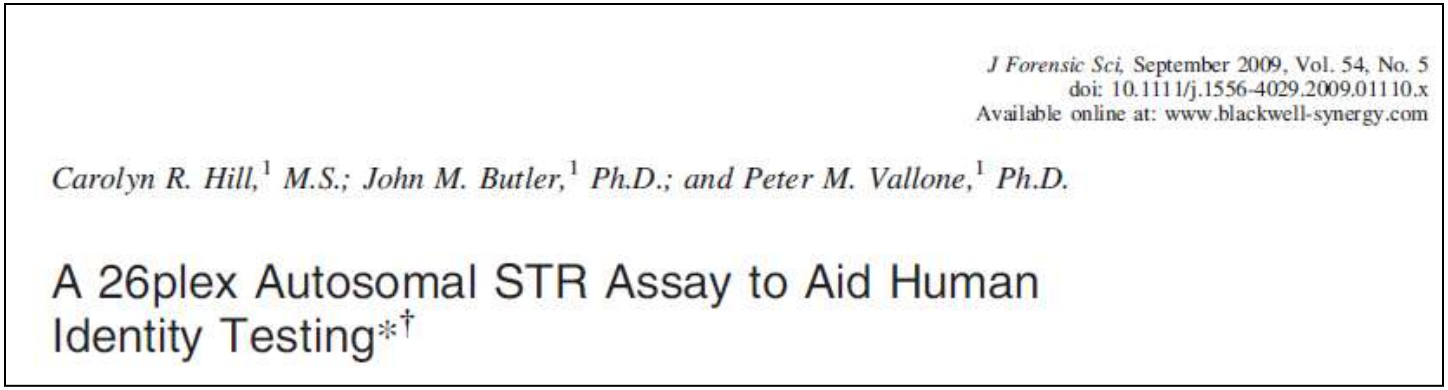
- What assay or kit data exist with 20plex (or greater) STR multiplexes?
  - NIST 26plex
  - PowerPlex 21 data collected at NIST

NIST 26plex Demonstration

- Our group at NIST has demonstrated that 25 autosomal STRs and amelogenin (26plex with 52 PCR primers) can be co-amplified with sensitivities similar to commercial STR kits

  – Hill, C.R., Butler, J.M., Vallone, P.M. (2009) A 26plex autosomal STR assay to aid human identity testing. *J. Forensic Sci.* 54(5): 1008-1015.

  – See also http://www.cstl.nist.gov/biotech/strbase/str26plex.htm



J Forensic Sci, September 2009, Vol. 54, No. 5
doi: 10.1111/j.1556-4029.2009.01110.x
Available online at: www.blackwell-synergy.com

Carolyn R. Hill,[1] M.S.; John M. Butler,[1] Ph.D.; and Peter M. Vallone,[1] Ph.D.

A 26plex Autosomal STR Assay to Aid Human Identity Testing*†

PowerPlex 21

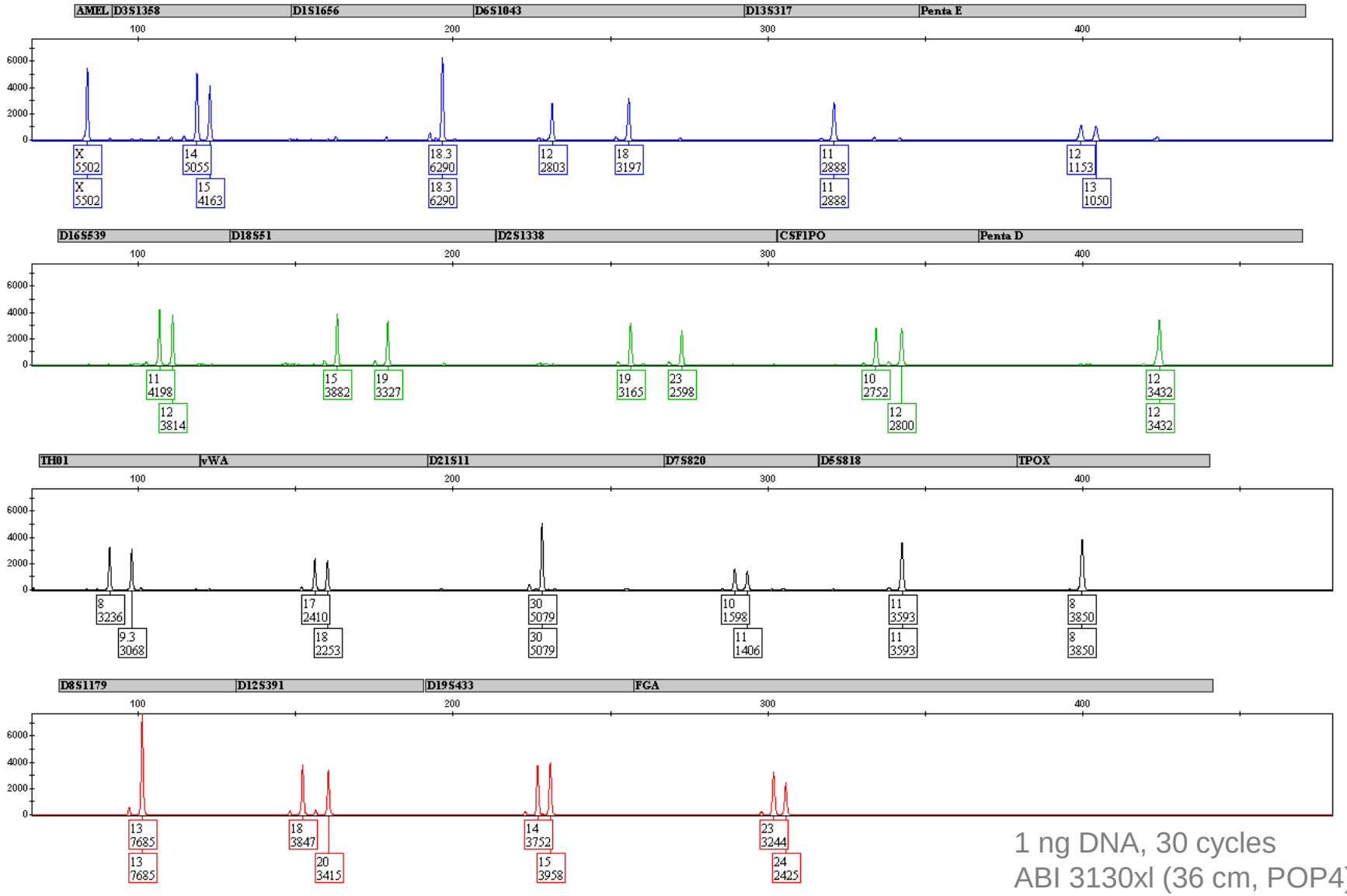
- Promega STR kit to be released in early 2012
  - NIST has been working with this kit since spring 2011 primarily for concordance testing and has permission from Promega to discuss results

- **Contains 20 autosomal STRs + amelogenin**

- **Enables examination of performance characteristics** similar to a future U.S. megaplex containing at least 20 loci

PowerPlex 21 NIST Result with 9947A

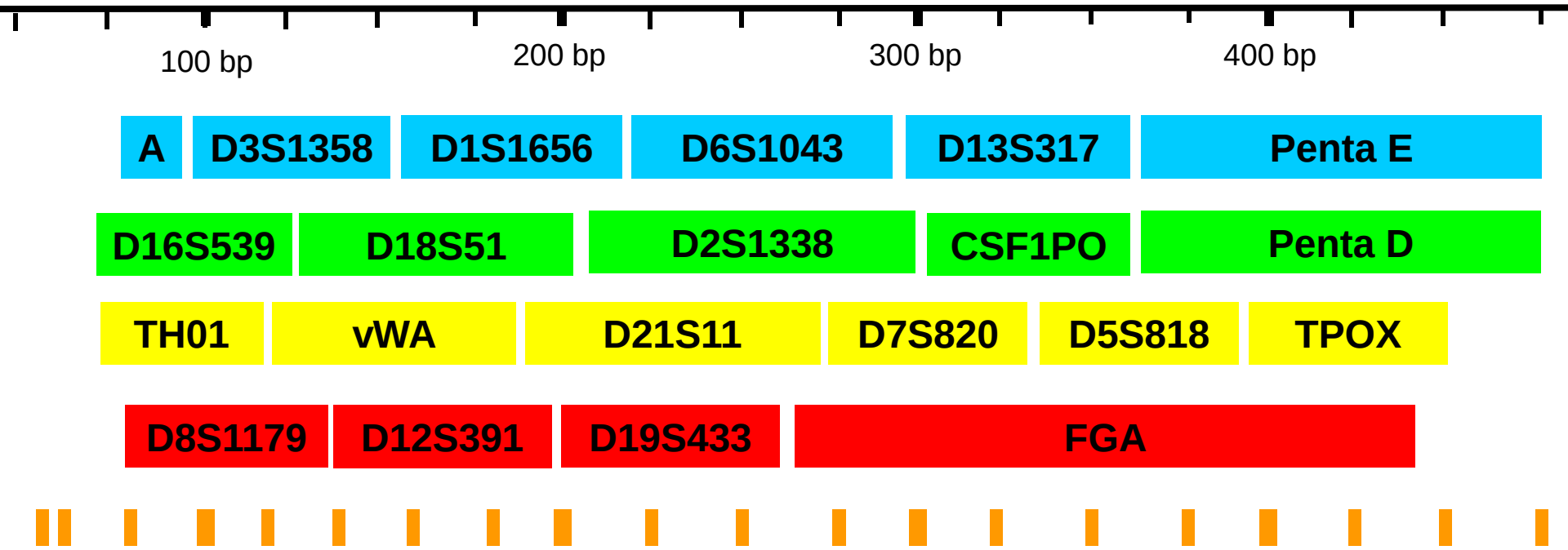
20 autosomal STR loci + amelogenin



1 ng DNA, 30 cycles
ABI 3130xl (36 cm, POP4)

PowerPlex 21

(to be released by Promega in early 2012)



Promega 5-dye kit

13 CODIS STRs + amelogenin

Penta D & Penta E (PP16 loci)

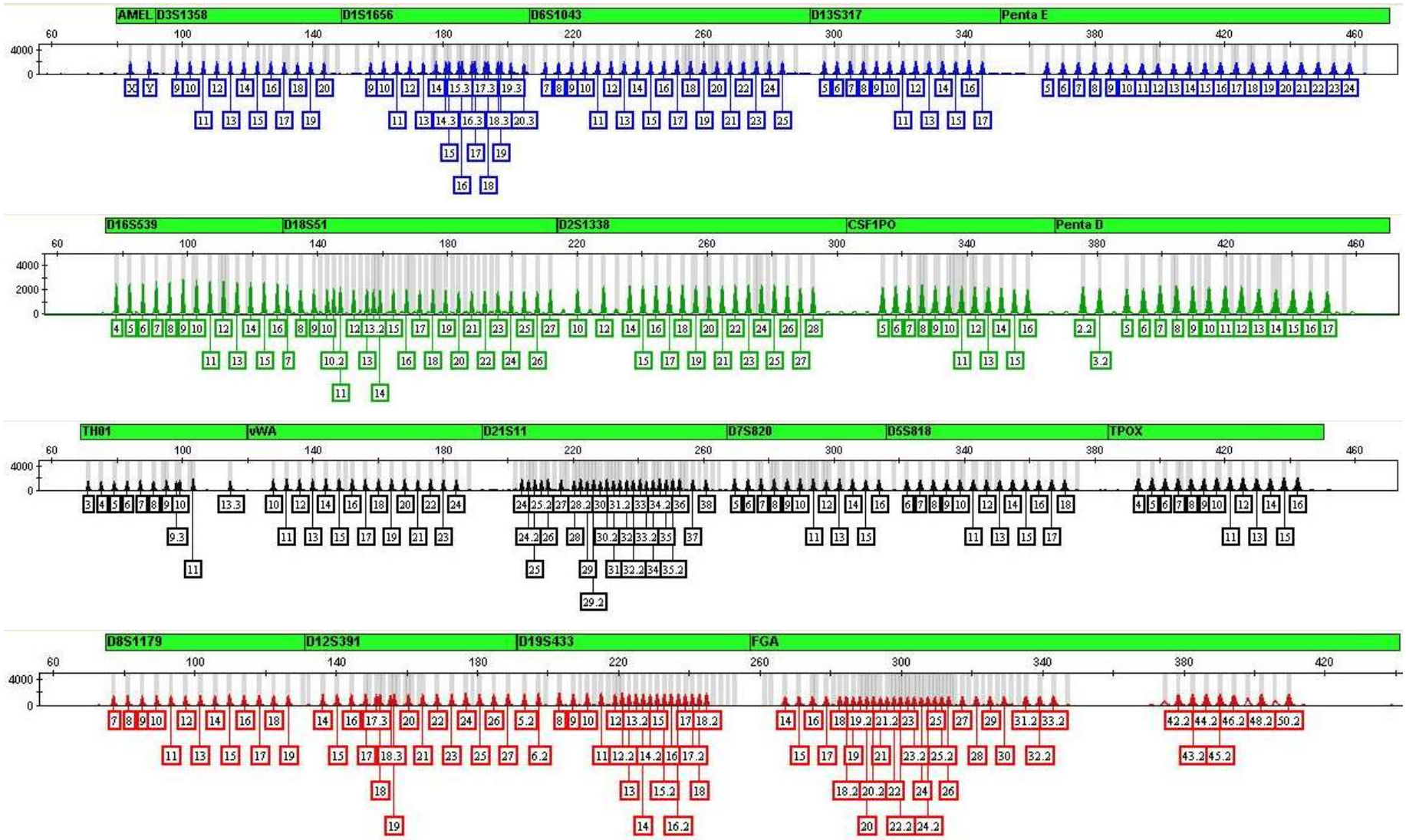
D2S1338 & D19S433 (Identifiler loci)

D12S391 & D1S1656 (best new European loci)

D6S1043 (previously only used in China – ABI Sinofiler kit – highly polymorphic)

PowerPlex 21 Allelic Ladders

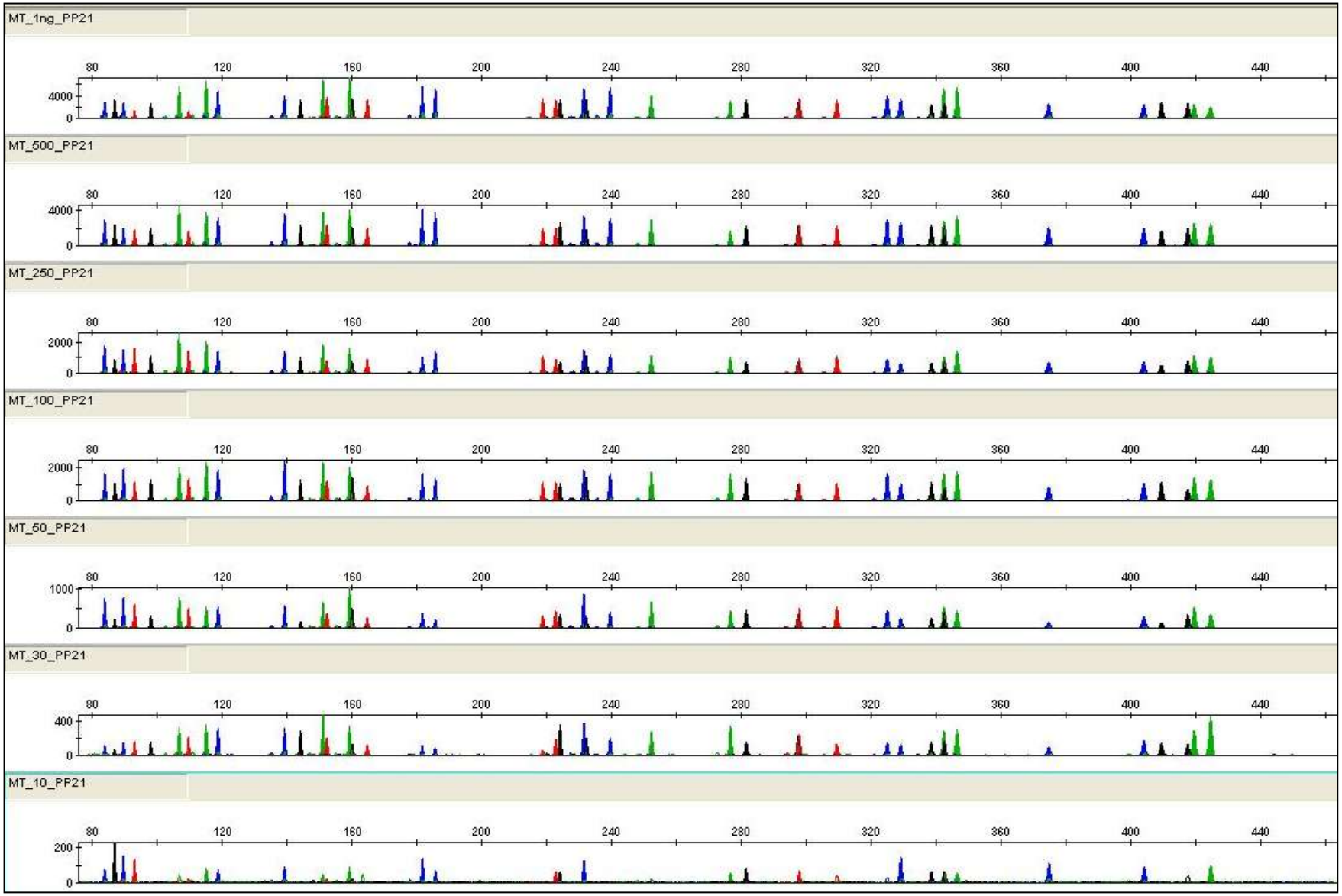
337 alleles present in this allelic ladder



Data courtesy of Becky Hill (NIST)

DNA Dilution Series with PowerPlex 21

As expected with any STR kit/assay, allele dropout occurs below 100 pg…



30 cycles, 3130xl data, NIST sample MT97150 (fully heterozygous)

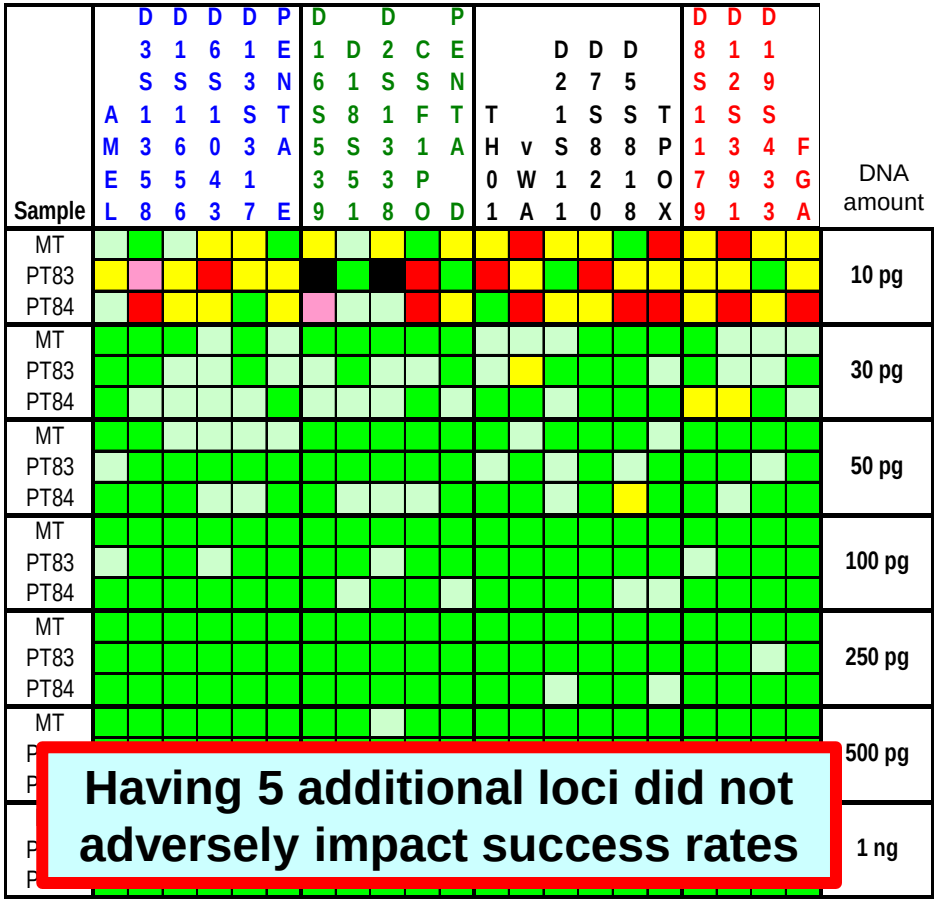
Data courtesy of Becky Hill (NIST)

Data courtesy of Becky Hill (NIST)

Measurement of Allele Dropout and Extreme Peak Height Imbalance for 2 STR Kits

Three fully heterozygous (except PT83 at Penta D) pristine DNA samples were examined in a dilution series with PowerPlex 21 and Identifiler Plus. Results are ordered by amplicon size and dye color.

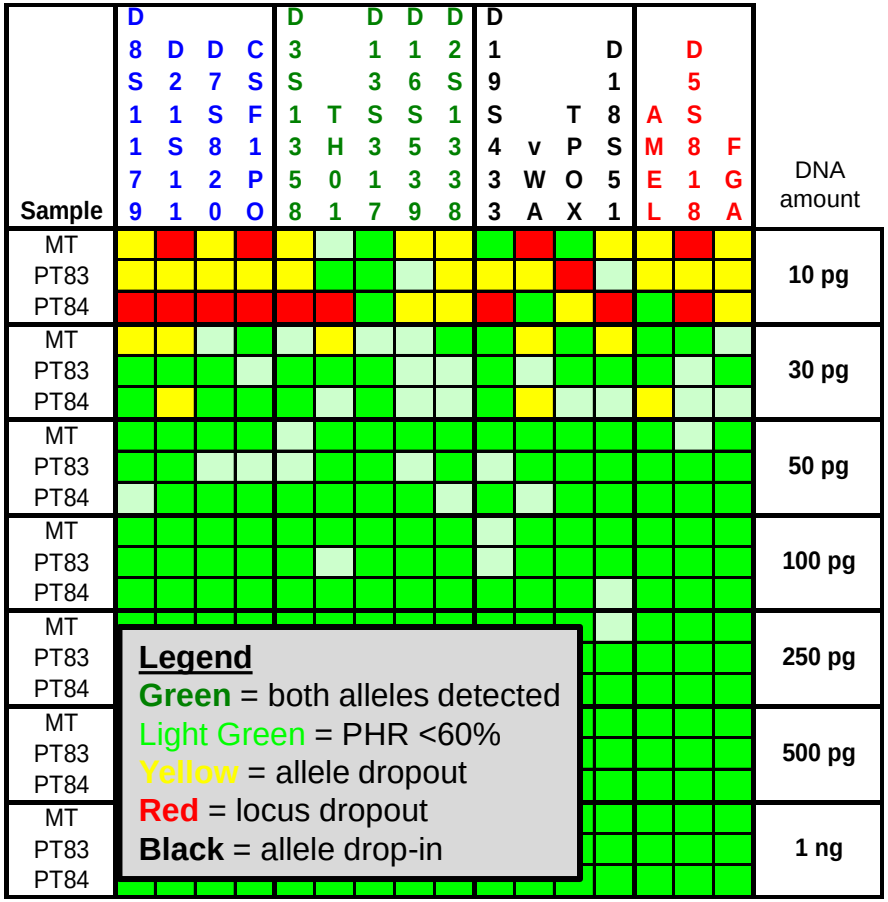
PowerPlex 21 - 30 cycles (5s@2kV)



Total alleles possible = 875
Total alleles present = 805

92% detected

Identifiler Plus - 28 cycles (10s@3kV)



Total alleles possible = 672
Total alleles present = 619

92% detected

Concern: DYS391 is a poor choice and AMEL Y nulls are not a significant concern

- AMEL Y nulls happen…
  - Common practice in some labs is a follow-up test with Y-STRs to confirm that a sample is male
  - **Some labs have implemented an additional ChrY test (SRY) to confirm AMEL Y nulls**

- A further purpose of having a single Y-STR is to aid QC checks if further Y-STR testing is performed for familial searching or casework purposes
  - DYS391 result will enable a QC check to Yfiler or PowerPlex Y results like D3 and D7 did for Profiler Plu/COfiler (albeit a rather weak one because it is not very polymorphic)
  - By itself, DYS391 is not polymorphic enough to be helpful with any potential familial search filter

SRY Male-Specific Amplicon Used
to Aid with Amelogenin Null Detection

Journal of Forensic Sciences, May 2009, 54(3): 551-555

J Forensic Sci, May 2009, Vol. 54, No. 3
doi: 10.1111/j.1556-4029.2009.01007.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Vanja Kastelic,[1] *B.S.; Bruce Budowle,*[2] *Ph.D.; and Katja Drobnič,*[1] *Ph.D.*

Validation of *SRY* Marker for Forensic Casework Analysis

[1]Forensic Science Centre, Ministry of the Interior, Vodovodna 95, Ljubljana, Slovenia.
[2]FBI Laboratory, Quantico, VA 22135.
Received 13 Mar. 2008; and in revised form 15 May 2008; accepted 14 July 2008.

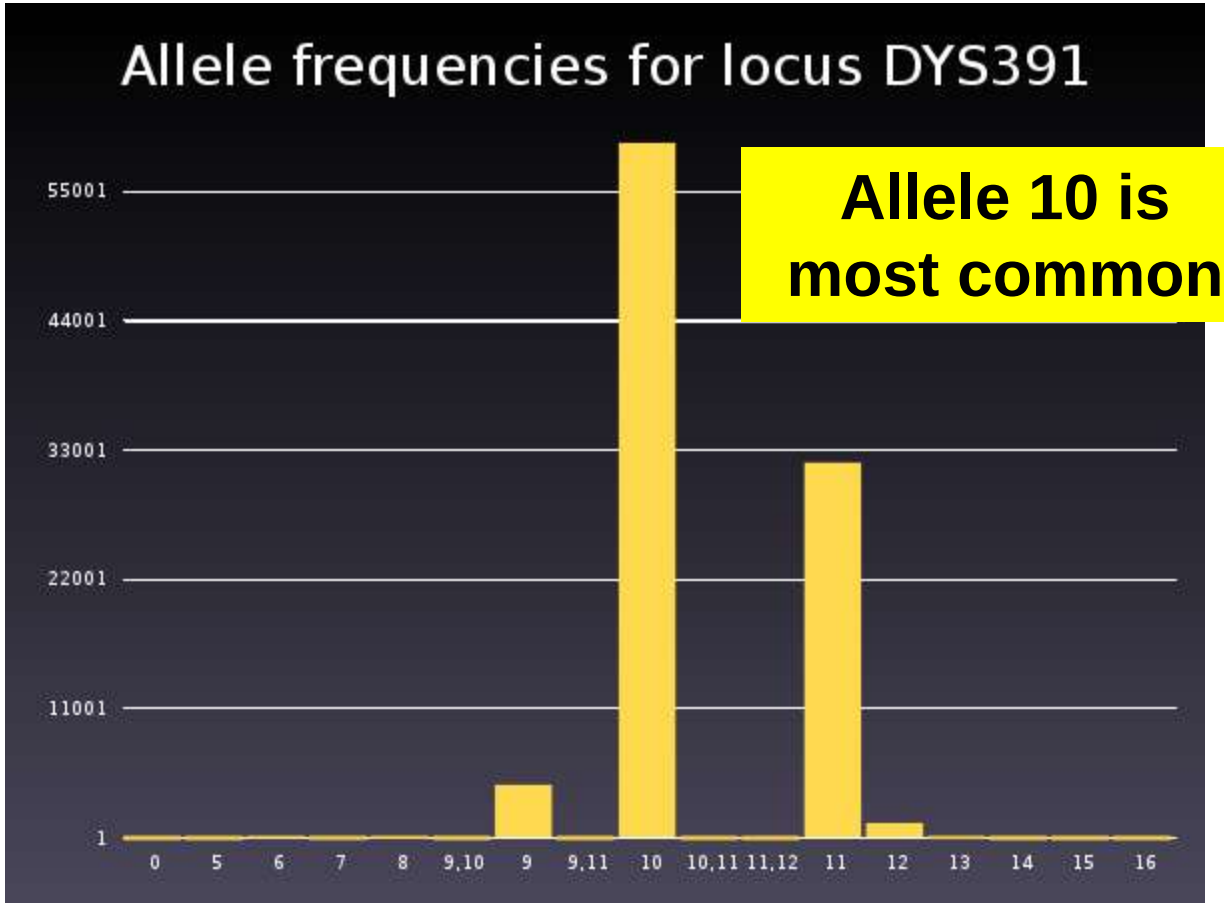
"Determining the gender of the source of forensic DNA evidence is based on the amelogenin test. However, **at times the assay may not be indicative of gender assignment, because of deletions at the amelogenin site.** …The study herein addresses the validation of primers for the target SRY gene regarding specificity, sensitivity, and robustness."

Why Consider DYS391?

- **DYS391 is located on the long arm of the Y-chromosome over 7 Mb away from amelogenin**. Thus, it is likely to be detected in the event of an amelogenin Y deletion that could make a male sample falsely appear as a female (X,-).
- DYS391 is not very polymorphic. From a data set of 97,575 haplotypes available on the Y-Chromosome Haplotype Reference Database, over half of them possess allele 10. However, only two null alleles have been reported and 0.01% duplication events (11 total) have been seen in over 700 different population groups from around the world. Thus**, it is a stable locus with a relatively narrow allele range**.
- DYS391 has a mutation rate of 0.26%, which is comparable to most autosomal STRs commonly in use. There have been 38 mutations observed so far in the 14,621 meioses reported in the literature and compiled on YHRD.

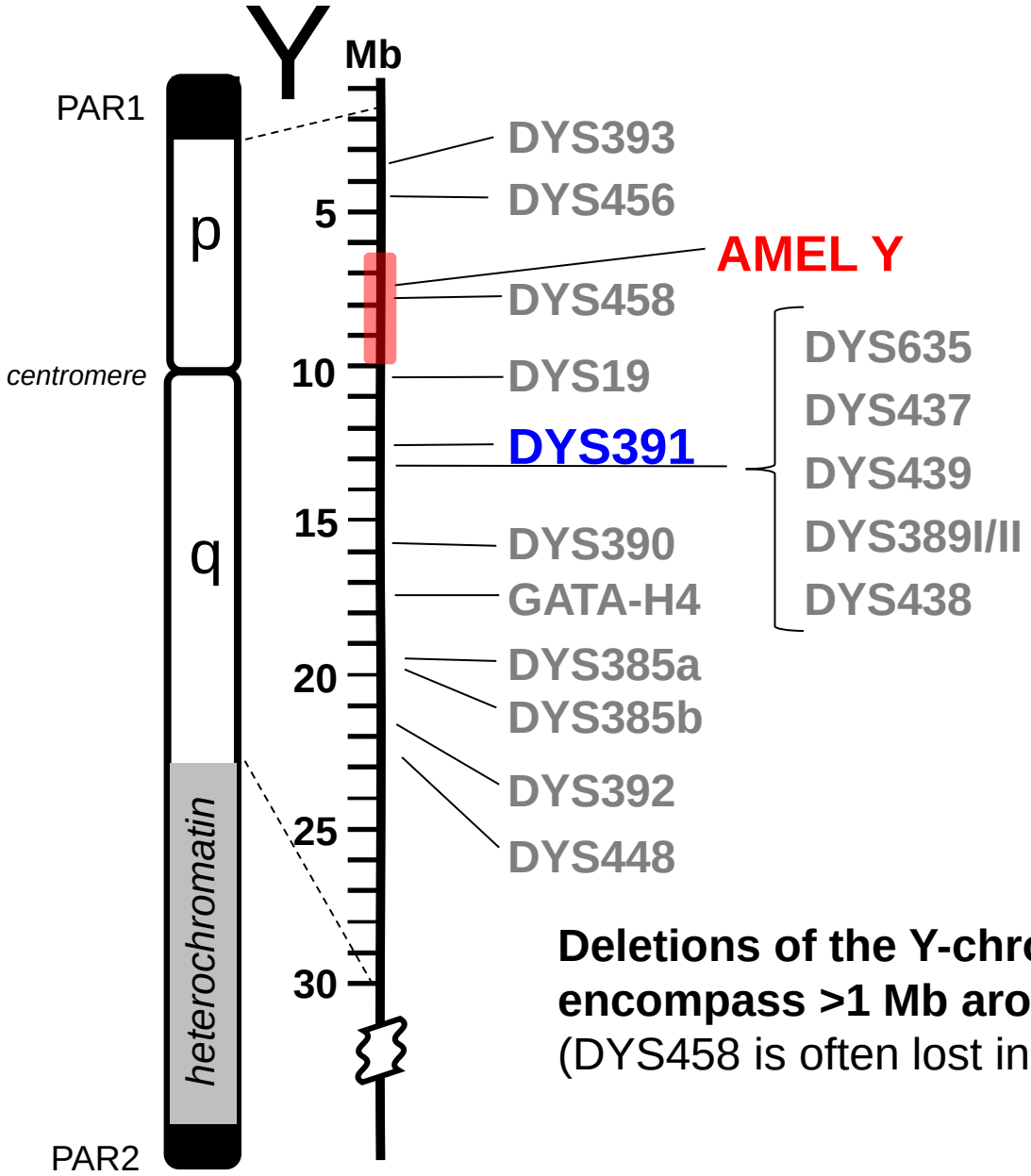
DYS391 Variability

YHRD (Y-chromosome Haplotype Reference Database) **data from 97,575 samples**



http://www.yhrd.org/Research/Loci/DYS391

Relative positions of 17 Y-STR loci commonly used in ChrY testing



Most commonly seen in males of Indian subcontinent origin

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

Deletions of the Y-chromosome can encompass >1 Mb around the AMEL Y region
(DYS458 is often lost in these situations)

Why mixing Y-STRs and autosomal STRs in a single DNA test is a bad idea…

Offender/arrestee reference samples

- Male samples: will work fine
- **Female samples: only autosomal STRs will amplify resulting in a waste of reagents compared to match probability produced**

Casework samples

- Mixtures: excess of either male or female DNA will result in poor STR typing results

Missing person samples

- Y-STRs will fail to work on female DNA samples

Do females represent a significant portion of the samples being examined in these specimen categories?

Not all DNA samples tested are male…

And if not male, then Y-STRs fail to amplify!

SDIS Offender/Arrestee Data:

- Virginia (371,000): ~**22% female***
- California (1.9 million): ~**17% female***
- Illinois (463,000): ~**16% female*** *Determined to be female based on amelogenin results or meta data

Data kindly provided by George Li & Brad Jenkins, Ken Konzak, and Taylor Scott

Missing Persons:

- **NamUs** (http://www.namus.gov/; searched 11/4/11):
  - Unidentified persons: **20% female** (1699/8438 cases)
  - Missing persons: **36% female** (3278/9012 cases)
- **NDIS Statistics** (Aug 2011):
  - Unidentified human remains: 5,324
  - Missing person cases: 1,039

Per NDIS Custodian (11/4/11):
~45% females in MP cases
(by amelogenin results)

Summary

- It is vital that an expanded set of core loci be carefully considered and implemented to avoid adventitious hits on large and growing DNA databases.

- There is limited “electrophoretic real-estate” in constructing STR multiplex assays that will work in 5-dye instruments and contain PCR products <500 bp – 6-dye kits and instruments will help.

- The number of females in DNA database and missing persons cases make required use of multiple Y-STRs of questionable utility.

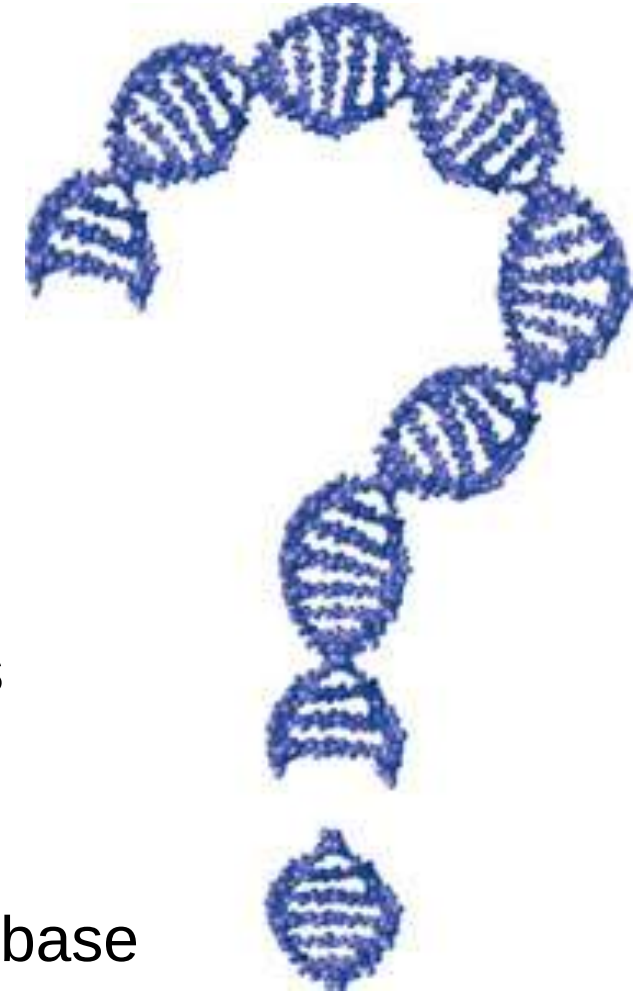
- **Data driven decisions are being made by the CODIS Core Loci Working Group.**

- The CODIS community will be involved in the implementation phase of adding new kits.

Thank you for your attention

Acknowledgments: Applied Biosystems, Promega, and Qiagen for STR kits used in concordance studies

CODIS Core Loci Working Group



Contact Information

John Butler
NIST Fellow
Group Leader of Applied Genetics
john.butler@nist.gov
301-975-4049

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

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