George Mason University

September 21, 2011 – Fairfax, VA

Forensic DNA Research

at the U.S. National Institute of Standards and Technology

Becky Hill

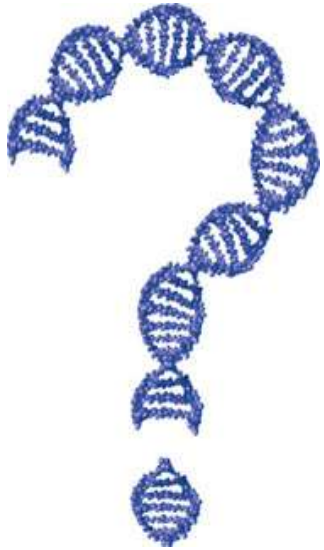
Applied Genetics Group

National Institute of Standards and Technology

Gaithersburg, Maryland

Presentation Outline

Please ask questions



- **A few things about me**
  - Background, interests, how did I get here?
- **NIST**
  - location, role, organizational structure, funding
- **Applied Genetics Group**
  - members, expertise, equipment, funding
- **Standard Reference Materials (SRMs)**
  - SRM 2391c: DNA Profiling Standard
- **Forensic DNA Research**
  - Concordance studies, miniSTRs and 26plex, low template DNA
- **Final thoughts and some advice for you…**

My Background

Introduction

- Have always been a math/science person
- Started off pre-med in my undergraduate studies
- Decided early on that research and development are REALLY where my interests lie
  - Held a few research positions in my undergrad experience (volunteer only)
- Graduated with a general Biology degree from University of Virginia

Professional Career

- Worked for 2 years at the American Red Cross, Jerome H. Holland Laboratory for Biomedical Research (Plasma Derivatives)
  - Worked on the Fibrin Sealant Bandage
- Worked for a biotech start-up company called Clearant for 5 years with many scientists from the Red Cross
  - Worked on the “Clearant Process” for hard and soft tissue allografts
- While at Clearant, I began my Masters Degree at George Mason University for Molecular Biology
  - Wrote my thesis on the “Clearant Process” and graduated in 2005

GMU = Forensics

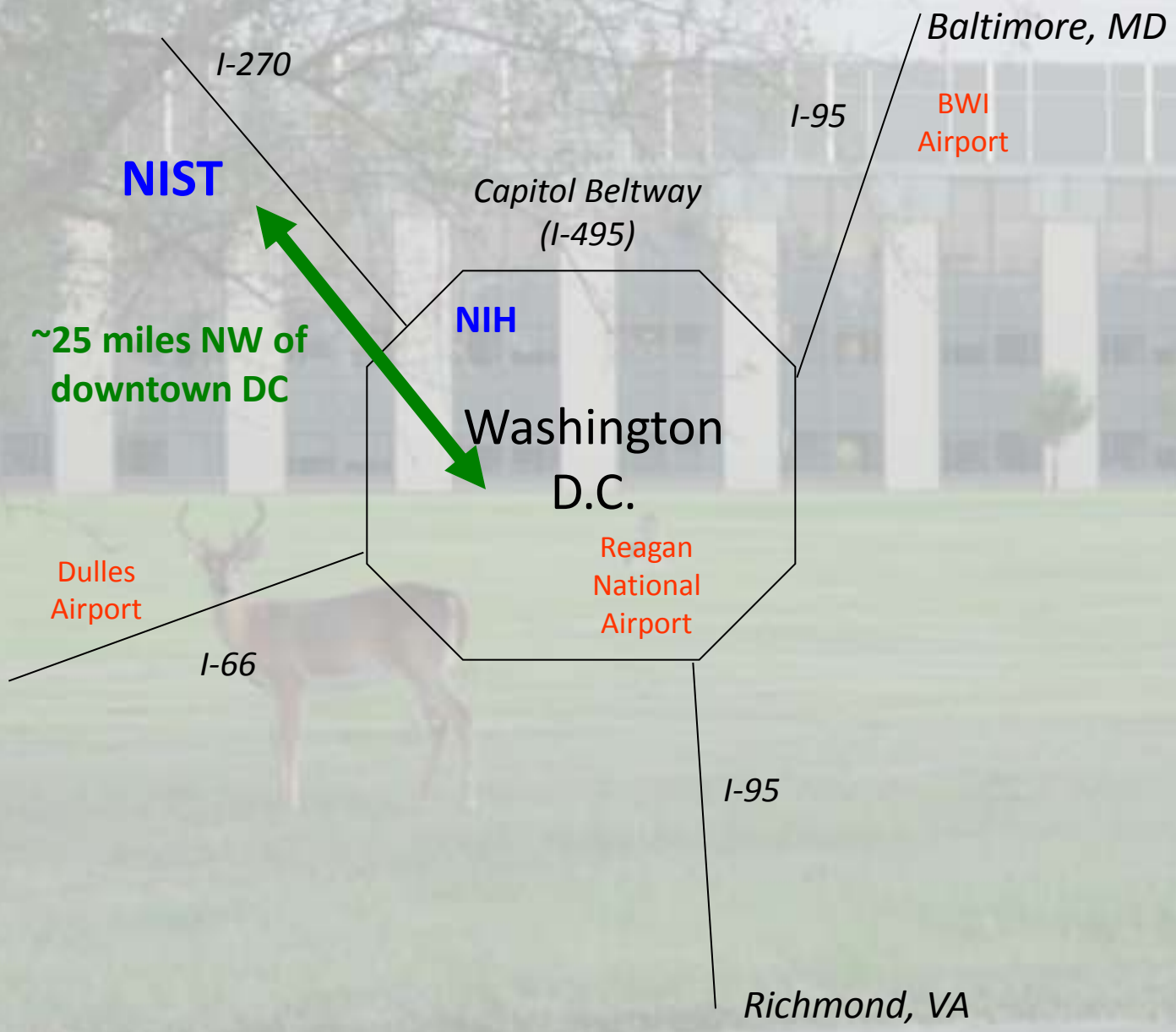
- Because of a forensic DNA course I took at GMU, I became VERY interested in this subject
  - Began to put my feelers out in the field for a job position
- Through a contact from Clearant, I heard about a job opening in the Human Identity Group at NIST
  - Interviewed and hired as an SAIC contractor (5 years)
  - Became federal government employee ~ 1 ½ ago
  - Been here for a total of 6 ½ years as a Forensic Biologist

NIST Background

U.S. National Institute of Standards and Technology

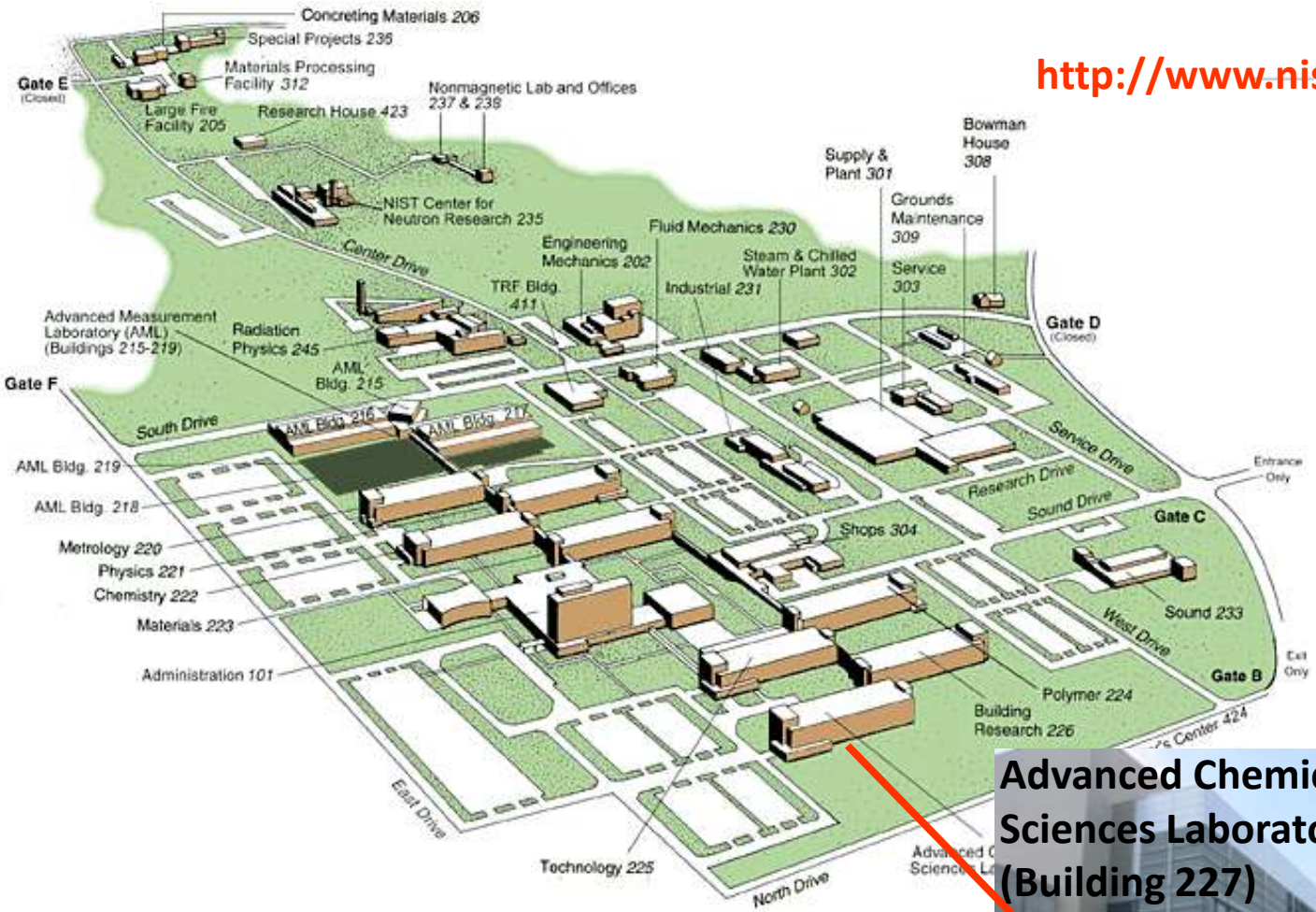
Department of Commerce

Location of NIST



NIST Gaithersburg Campus

http://www.nist.gov



National Institute of Standards & Technology (NIST)

- **Non-regulatory agency** established in 1901 in the US Department of Commerce.
- Mission to promote US innovation and industrial competitiveness by advancing measurement science, standards & technology.
- NIST is at the top of the US standards pyramid for a wide variety of physical standards, test methods, and calibrations.

Early Driver for U.S. Standards

1904

- Out-of-town fire companies arriving at a Baltimore fire cannot couple their hoses to the hydrants. 1526 buildings razed.

1905

- National Fire Protection Association adopted NBS-developed national hose coupling standard.

NIST Today

Major Assets

- ~ 2,900 employees
- ~ 2600 associates and facilities users
- ~ 400 NIST staff on about 1,000 national and international standards committees
- 3 Nobel Prizes in past 15 years

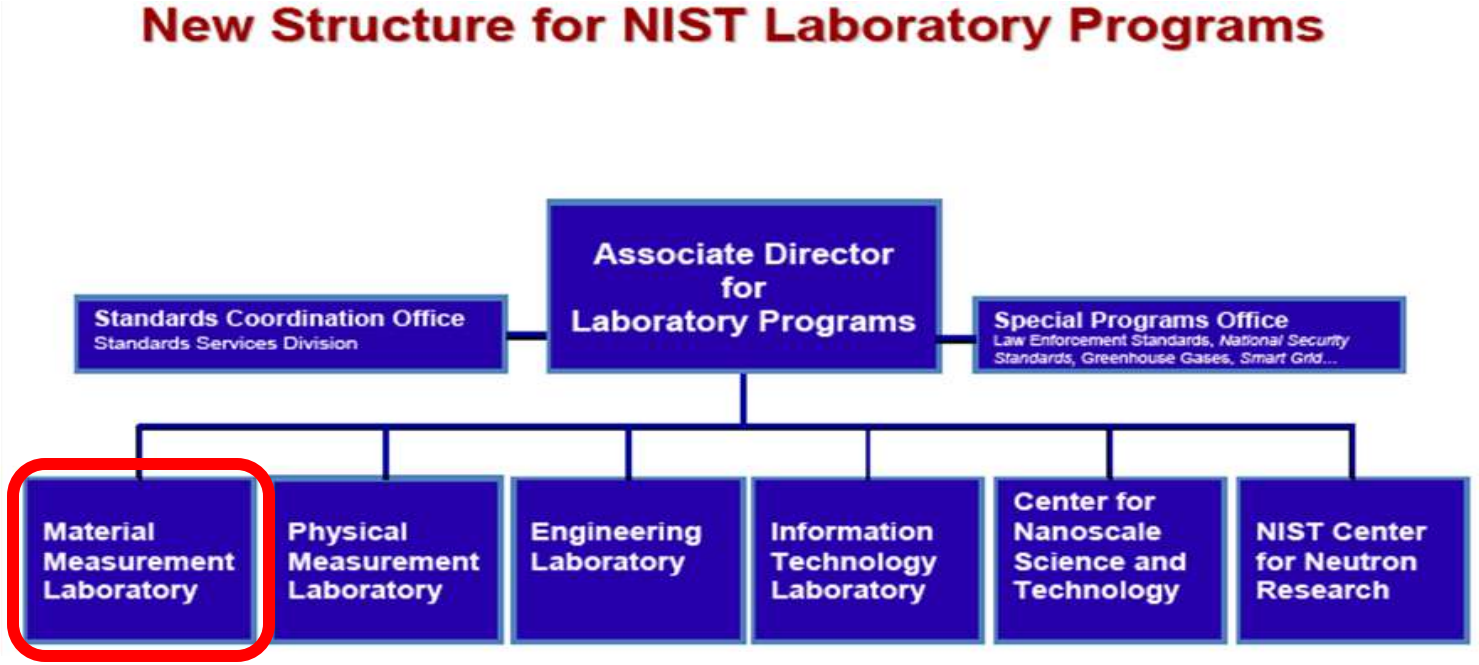
Major Programs

- **NIST Laboratories**
- Baldridge National Quality Program
- Hollings Manufacturing Extension Partnership
- Technology Innovation Program

Joint NIST/University Institutes:

- JILA
- Joint Quantum Institute
- Institute for Bioscience & Biotechnology Research
- Hollings Marine Laboratory

The NIST Laboratories



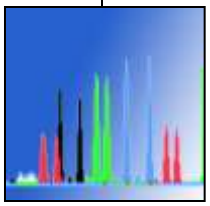
Traditionally focused research and measurement service activities on physical science and engineering disciplines

Bioscience and Health identified as a new area for significant emphasis for NIST labs

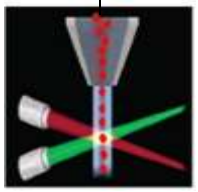
NIST Biochemical Science Division

Laurie Locascio
Division Chief

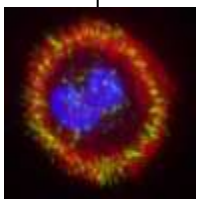
Process Sensing
Dean Ripple
Group Leader



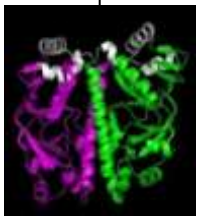
Applied Genetics
John Butler
Group Leader



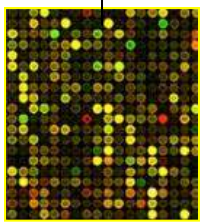
Bioassay Methods
Steve Choquette
Group Leader



Cell Systems Science
Anne Plant
Group Leader



Macromolecular Structure & Function
John Marino
Group Leader



Multiplexed Biomolecular Science
Marc Salit
Group Leader

NIST Applied Genetics Group

NIST
Applied
Genetics

Applied Genetics
Group Mission Statement

Advancing technology and traceability through quality genetic measurements to aid work in

- **forensic DNA testing**
- **clinical diagnostics**
- **cell line authentication**
- **agricultural biotechnology**
- **DNA biometrics**

NIST Applied Genetics

APPLIED GENETICS Group

Major Programs Currently Underway

- **Forensic DNA**
  - New loci and assays (26plex)
  - STR kit testing
  - Ancestry SNP assays
  - Low-template DNA studies
  - Mixture interpretation
  - STR nomenclature
  - Variant allele cataloging and sequencing
  - Expert systems review
  - Training workshops to forensic DNA laboratories
  - Validation information and software tools
  - Textbook – 3rd ed. (2 vol.)
- **Clinical Genetics**
  - **Huntington's Disease SRM**
  - **CMV SRM**
  - **Exploring future needs**
- **Ag Biotech**
  - "universal" GMO detection/ quantitation (35S promoter)
- **DNA Biometrics**
  - Rapid PCR methods
  - Efforts to standardize testing of future portable DNA systems
  - Kinship analysis
- **Cell Line Authentication**

NIST Applied Genetics

Group Expertise and Funding Sources

Group Expertise

- Reference Material Characterization
- Standard Information Resource Development
- Rapid Multiplex PCR Assay Construction
- Short Tandem Repeat (STR) Genotyping
- Single Nucleotide Polymorphism (SNP) Genotyping
- DNA Sequencing
- Training Materials and Workshops (validation info)

Current Funding Sources

- **National Institute of Justice** (Forensic DNA)
- **FBI Science & Technology Branch** (DNA Biometrics)
- **NIST SRM Program** (SRM development and production)
- **Base funding from Congress (clinical DNA)**

Applied Genetics Group Instrumentation

- **ABI 3130xl and 3500** for Sanger sequencing, SNP analysis, and STR genotyping
- **ABI 7500** for qPCR (DNA quantitation)
- **ABI 9700** and Veriti thermal cyclers for PCR
- **Fluidigm BioMark** for digital PCR (copy number determination)

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

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

mtDNA & Mass Spec

NIJ National Institute of Justice

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

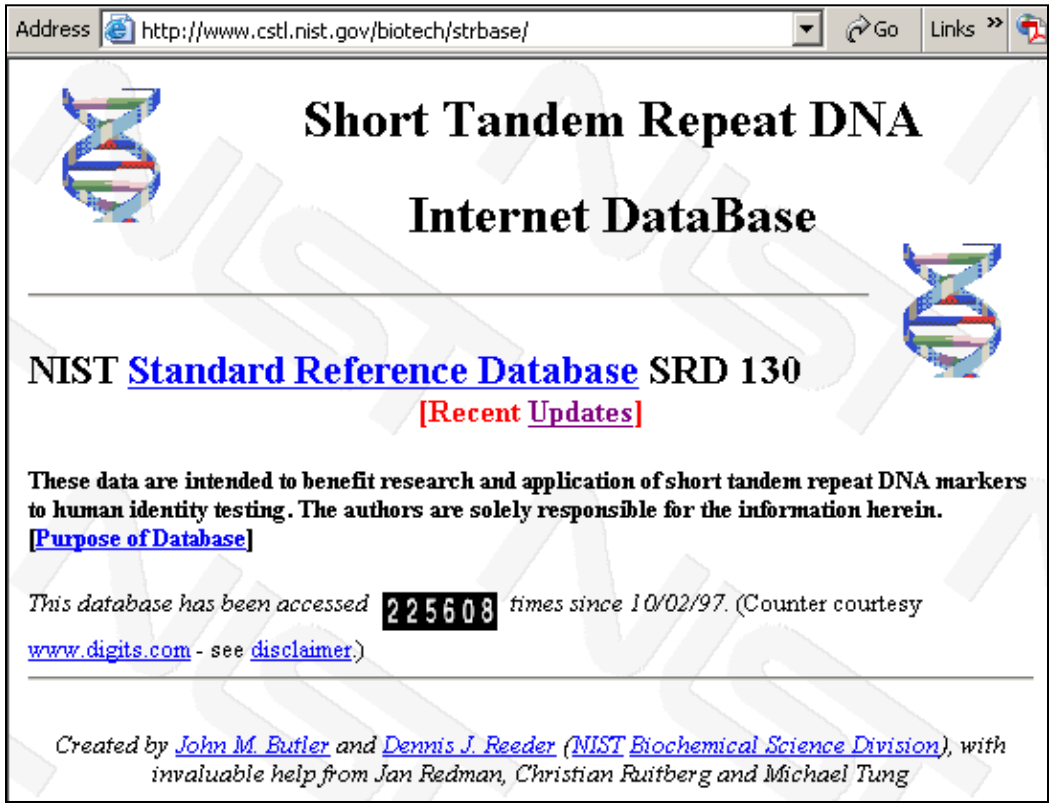
Current Areas of NIST Effort with Forensic DNA

- **Standards**
  - Standard Reference Materials
  - Standard Information Resources (STRBase website)
  - Interlaboratory Studies
- **Technology**
  - Research programs in STRs, SNPs, miniSTRs, Y-STRs, mtDNA, qPCR, LT-DNA, mixtures, rapid PCR
  - Assay and software development, expert system review
- **Training Materials**
  - Textbooks, review articles and workshops on STRs, CE, validation
  - PowerPoint and pdf files available for download
  - Training workshops conducted to scientists, lawyers, and students

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

Information Resource

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



Includes information on:

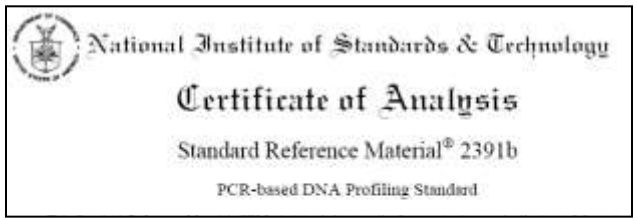
- Core STR loci
- Validation
- STR reference list
- NIST publications
- miniSTRs
- Forensic SNPs
- Variant STR alleles
- Population data resources
- Addresses of scientists

Provides up-to-date information and has been used in court cases to support application of DNA technology

Standard Reference Materials (SRMs)

http://www.nist.gov/srm

Traceable standards to ensure accurate and comparable measurements between laboratories

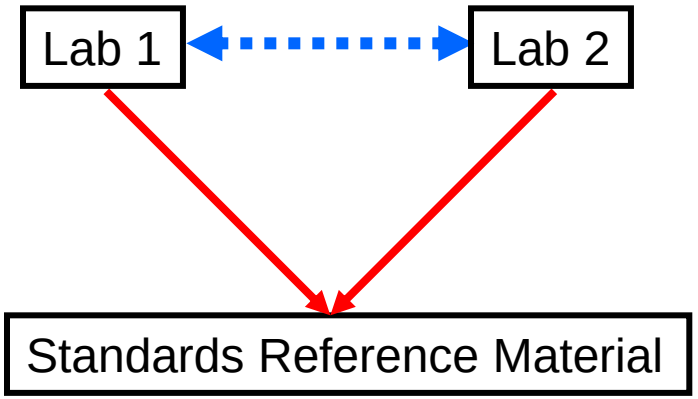

National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 2391b

PCR-based DNA Profiling Standard

SRM 2391b – autosomal STRs
SRM 2392 &-I – mtDNA sequencing
SRM 2395 – Y-STRs
SRM 2372 – DNA quantitation
SRM 2394 – mtDNA heteroplasmy
SRM 2399 – Fragile X



Calibration with SRMs enables confidence in comparisons of results between laboratories

Helps meet ISO 17025 needs
for traceability to a national metrology institute

2003: NIST SRM 2391b

Driven primarily by commercial kit loci…

National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 2391b

PCR-based DNA Profiling Standard

This Standard Reference Material (SRM) is intended primarily for use in the standardization of forensic and paternity quality assurance procedures for Polymerase Chain Reaction (PCR)-based genetic testing and for instructional law enforcement or non-clinical research purposes. This SRM can also be used for quality assurance when assigning values to in-house control materials. It is not intended for any human or animal clinical diagnostic use. Note that SRM 2391b is slightly modified from SRM 2391, in that there is more emphasis on Short Tandem Repeats (STRs) and less emphasis on D1S80 [1,2] reflecting the growing interest and utility of STRs [3 to 14]. Additional information on each STR locus can be found at a NIST-sponsored database on the internet: http://www.cstl.nist.gov/biotech/strbase [14].

This SRM is composed of well-characterized human deoxyribonucleic acid (DNA) in two forms: genomic DNA and DNA to be extracted from cells spotted onto filter paper. A unit of the SRM is composed of 12 frozen components packaged in one box. See the section in this certificate entitled *Description of Components* for a complete listing of the components.

Certified Values: The SRM is certified for genetic loci of forensic interest that were commercially available at the time of production. Genetic types for these loci can be found in Tables 1, 2, and 3. The tables are organized as follows: Table 1 lists the genetic types for the Federal Bureau of Investigation's (FBI's) CODIS (COmbined DNA Index System) core STR loci; Table 2 lists additional STR loci of interest; and Table 3 lists the genetic types for D1S80, AmpliType® PM + HLADQA1, and Amelogenin.

Expiration of Certification: The certification of this SRM is valid until **31 December 2008**, provided the SRM is handled and stored in accordance with the instructions given in this certificate. However, the certification is invalid if the SRM is contaminated or otherwise modified.

Maintenance of SRM Certification: NIST will monitor this SRM over the period of its certification. If substantive technical changes occur that affect the certification before the expiration of certification, NIST will notify the purchaser. Return of the attached registration card will facilitate notification.

Storage: Store frozen at a temperature of -20 °C. **DO NOT** use a self-defrosting freezer because periodic cycling of temperatures may cause shortened shelf life of this SRM.

The overall direction and coordination of the technical activities leading to certification were under the chairmanship of J.M. Butler of the NIST Biotechnology Division.

Analytical determination and technical measurements leading to the certification of this SRM were performed by M.C. Kline and J.W. Redman of the NIST Biotechnology Division.

The support aspects involved in the preparation, certification, and issuance of this SRM were coordinated through the NIST Standard Reference Materials Group by C.S. Davis.

Vincent Vilker, Acting Chief
Biotechnology Division

Gaithersburg, MD 20899
Certificate Issue Date: 06 December 2002

John Rumble, Jr., Chief
Measurement Services Division

SRM 2391b

Page 1 of 7

2. Certified Values for Additional STR Loci

F13B	FES/FPS	LPL	Penta D	Penta E	D2S1338	D19S433
10,10	12,12	10,11	10,15	7,12	17,23	13,16.2
8,10	10,11					16
9,10	11,12					4
6,9	10,13					3
8,9	11,13					14
9,10	11,11	10,12	9,12	12,14	25,25	12,14
6,8	11,11*	11,12	3.2,11	12,16	17,22	13,15.2
6,8	10,11	9,11	8,9	5,10	22,22	12.2,15
8,10	10,12	11,12	12,12	12,13	19,23	14,15
8,8	11,11	10,12	8,12	11,11	23,23	13,14
8,10	10,12	11,12	12,12	12,13	19,23	14,15
8,8	11,11	10,12	8,12	11,11	23,23	13,14

48 autosomal STRs characterized across 12 DNA samples

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 (price reduction)

STR Kits Tested with SRM 2391c

Kit Provider			Primer Mixes
Life Technologies	*Promega*	*Qiagen*	*NIST*
Identifiler	Powerplex 16	ESSplex	26plex [3]
Identifiler Plus	Powerplex 16 HS	IDplex	miniSTRs [4,5]
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			

Alleles sequenced:
SE33
D12S391
D1S1656
Penta D
Penta E
D8S1115

22 commercial STR kits examined

NIST developed 26plex and miniplexes

No discordant results observed on SRM 2391c samples

Forensic DNA Research Programs

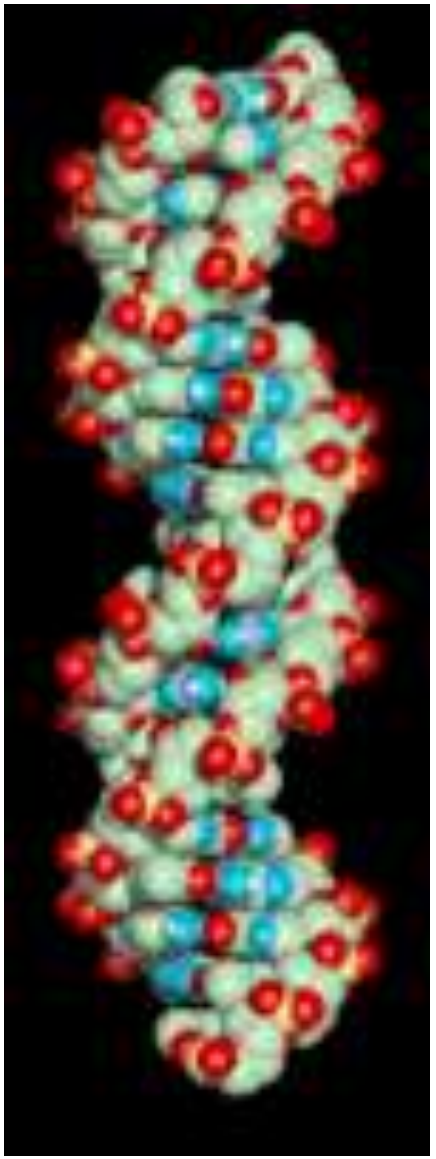
Concordance Studies

miniSTRs and the 26plex

Low template DNA

Methods for Human Identification

Fingerprints have been used since 1901



DNA since 1986

Steps in DNA Analysis

Usually 1-2 day process (a minimum of ~5 hours)

Steps Involved

- Collection
- Specimen Storage
- Extraction
- Quantitation
- Multiplex PCR
- STR Typing
- Interpretation of Results
- Database Storage & Searching
- Calculation of Match Probability

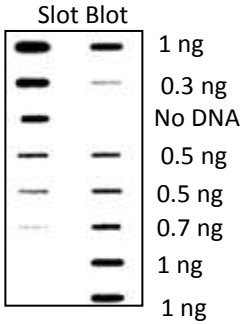
Blood Stain

Buccal swab

Sample Collection & Storage

Biology

DNA Extraction



DNA Quantitation

Multiplex PCR Amplification

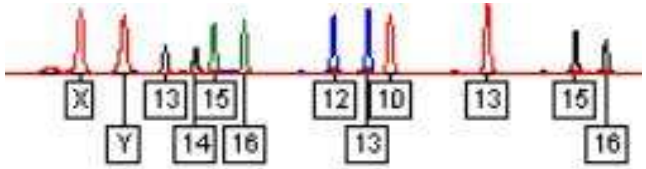
Genetics

If a match occurs, comparison of DNA profile to population allele frequencies to generate a case report with probability of a random match to an unrelated individual

DNA Database Search

Technology

DNA separation and sizing



STR Typing

Male: 13,14-15,16-12,13-10,13-15,16

Interpretation of Results

Basis of DNA Profiling

The genome of **each individual is unique** (with the exception of identical twins) and **is inherited from parents**

Probe subsets of genetic variation in order to differentiate between individuals (statistical probabilities of a random match are used)

DNA typing must be **performed efficiently and reproducibly** (information must hold up in court)

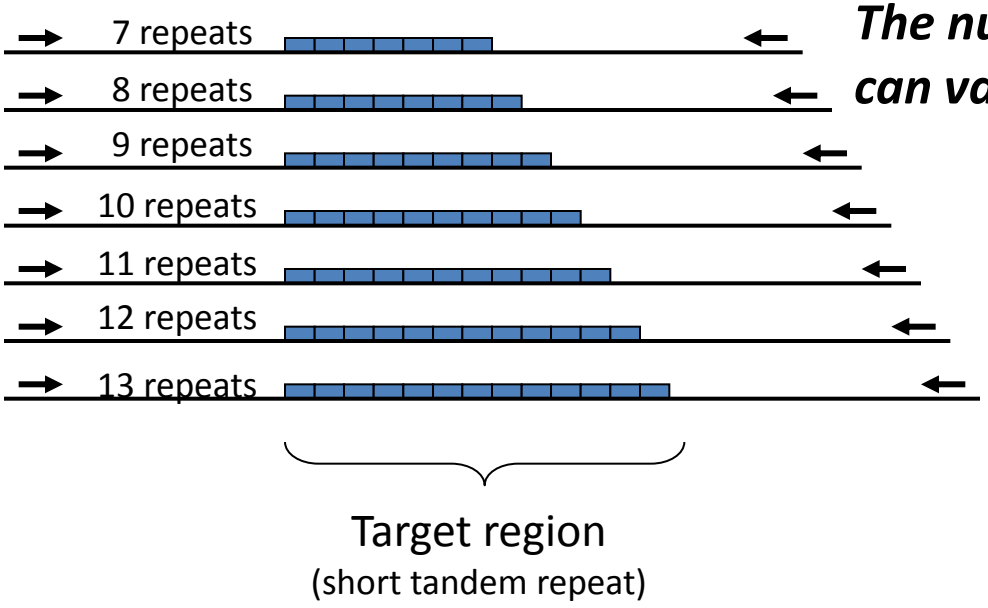
Current standard DNA tests **DO NOT look at genes** – little/no information about race, predisposal to disease, or phenotypical information (eye color, height, hair color) is obtained

Short Tandem Repeat (STR) Markers

An accordion-like DNA sequence that occurs between genes

TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGAAGACAGGTG**GATAGATA**
GATAGATAGATAGATAGATAGATAGATAGATAGATAGATATCATTGAAAGACAAA
ACAGAGATGGATGATAGATACATGCTTACAGATGCACAC

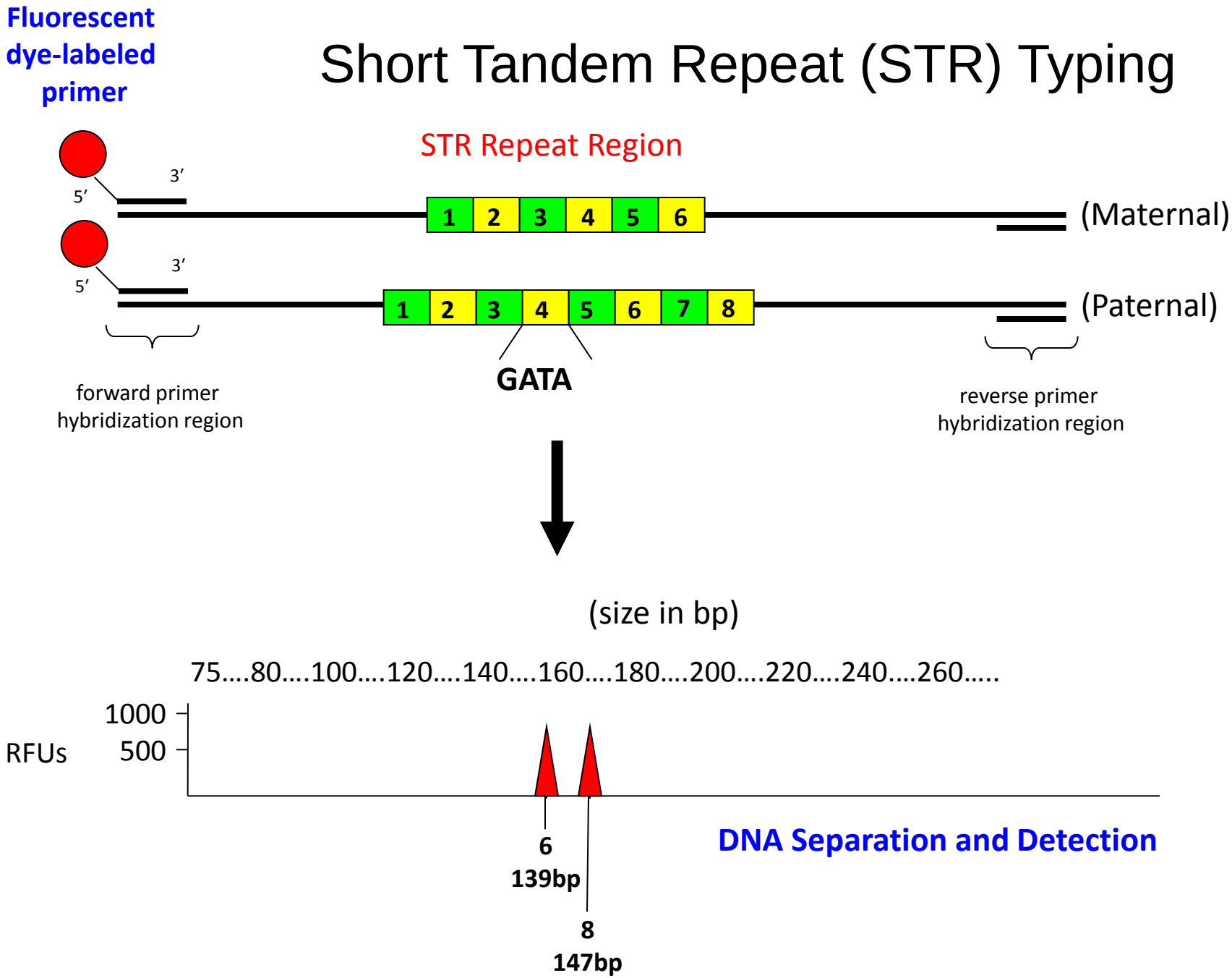
= 12 GATA repeats ("12" is all that is reported)



The number of consecutive repeat units can vary between people

The FBI has selected **13 core STR loci** that must be run in all DNA tests in order to provide a common currency with DNA profiles

Short Tandem Repeat (STR) Typing

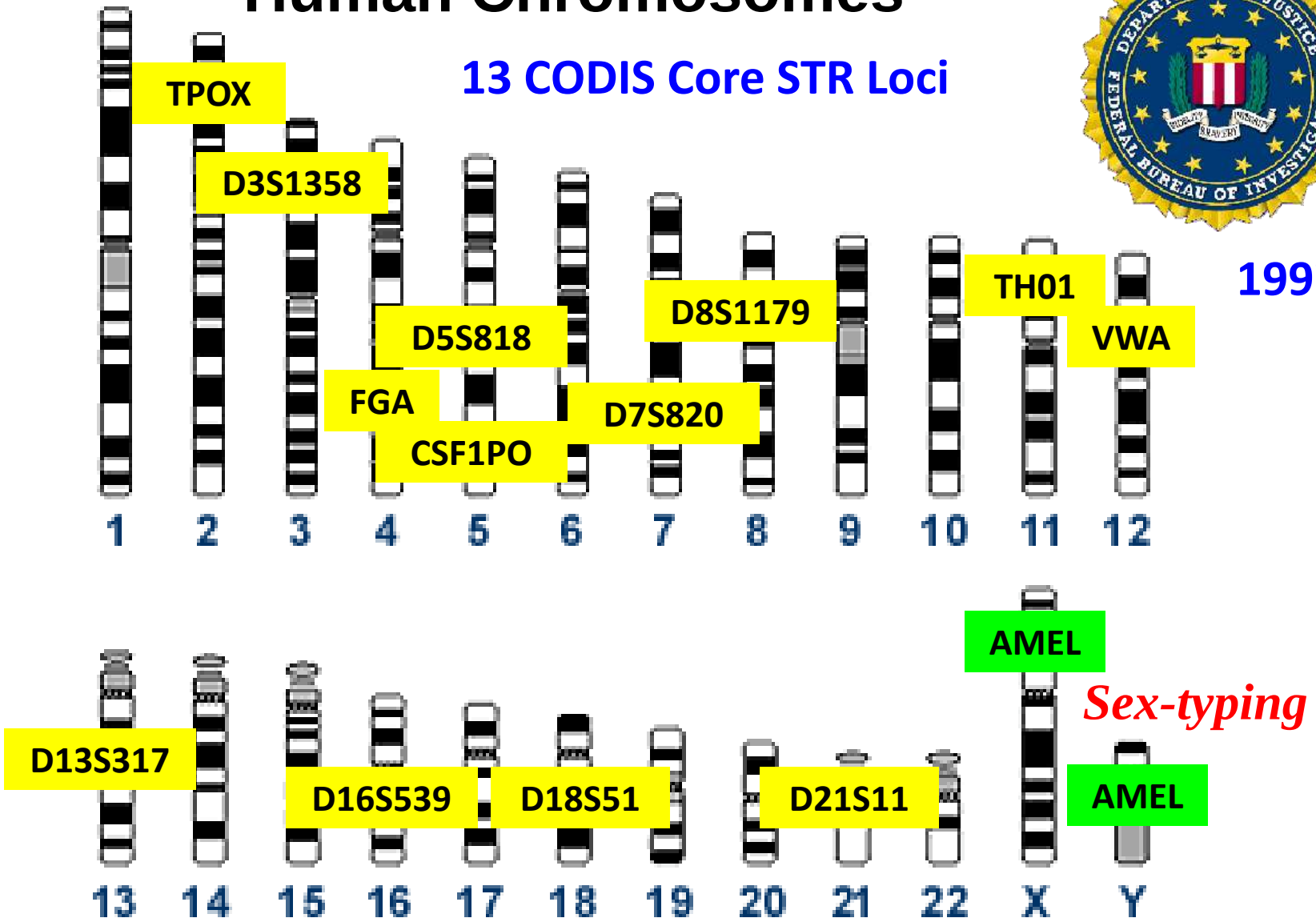


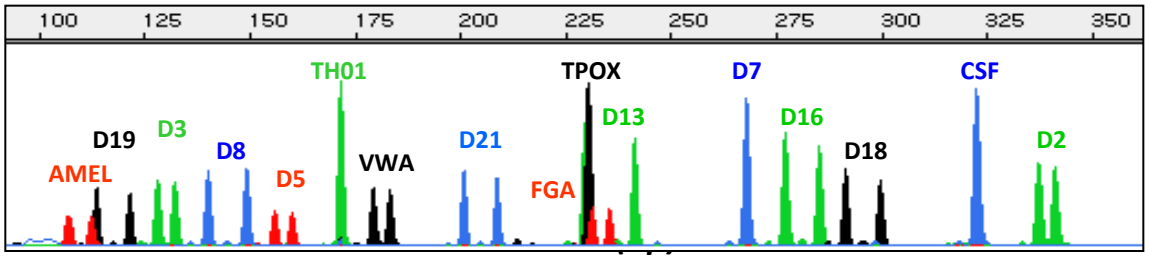
Position of Forensic STR Markers on Human Chromosomes

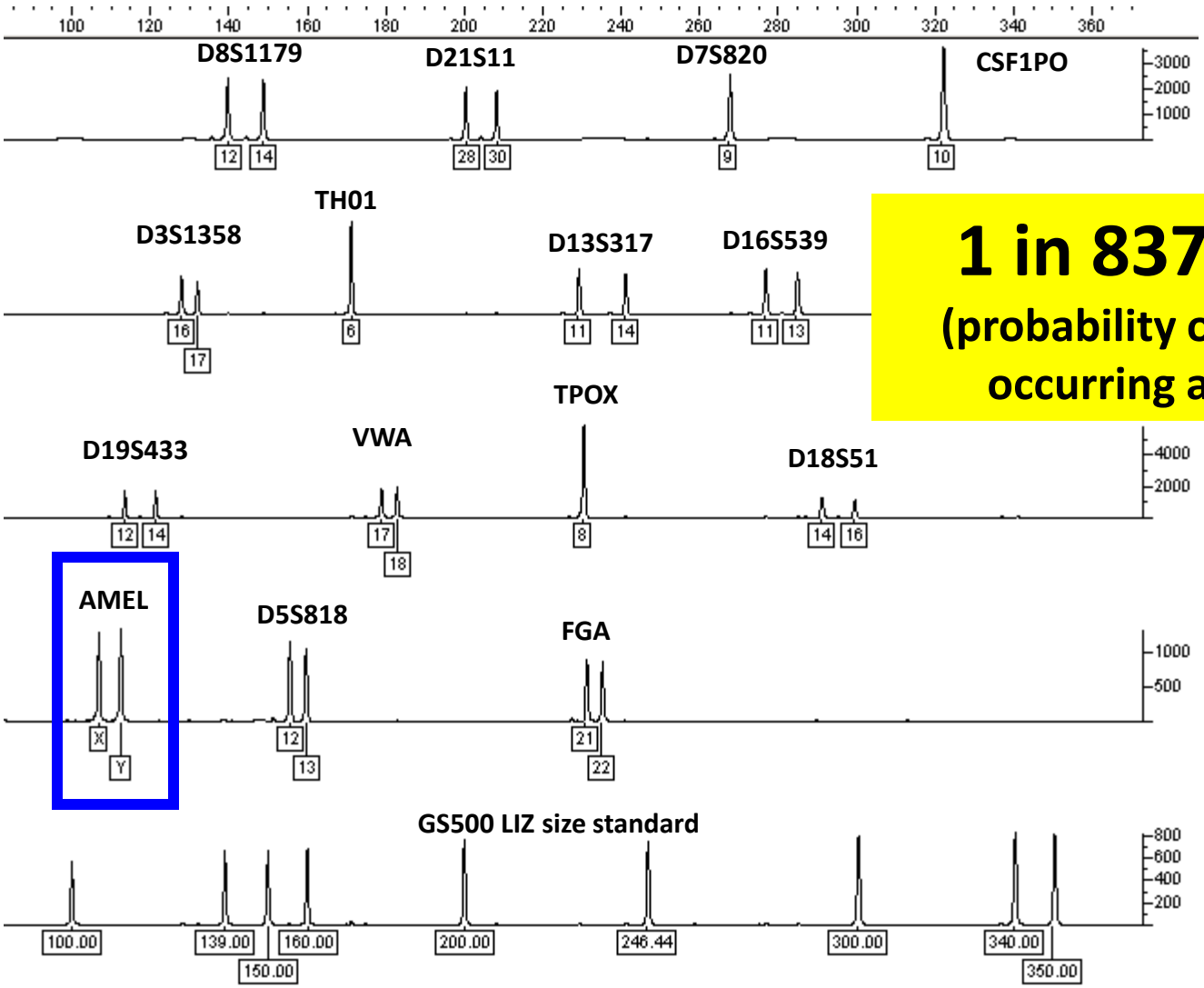
Core STR Loci for the United States

13 CODIS Core STR Loci

1997







1 in 837 trillion

(probability of this profile occurring at random)

NIST Pipeline for STR Kit Analysis

- Concordance testing with standard samples
  - Sequence analysis of any null alleles to understand differences
- Locus characteristics
  - Heterozygote peak height ratios
  - Stutter percentages (including allele-specific)
- Allele frequencies for all new loci
  - Across U.S. Caucasian, Hispanic, African American, and Asian
- Probability of identity for different locus sets

Characterization of New STR Loci

- **23 loci now present in commercial STR kits**
  - 13 CODIS loci plus **D2S1338** (40 alleles), **D19S433** (36 alleles), **Penta D** (50 alleles), **Penta E** (54 alleles), **D2S441** (22 alleles), **D10S1248** (13 alleles), **D22S1045** (14 alleles), **D12S391** (51 alleles), **D1S1656** (25 alleles), and **SE33** (171 alleles)

- Chromosomal location
- Repeat structure and sequence
- U.S. population samples
- Literature surveys to gather all known alleles

23 STR loci present in STR kits

STR Locus	Alleles Observed	Genotypes Observed	H(obs)	P_I (all samples) n = 1426
SE33	58	341	0.9383	0.0063
Penta E*	20	113	0.8779	0.0175
D2S1338	13	73	0.8752	0.0221
D1S1656	17	99	0.8871	0.0229
D18S51	23	102	0.8696	0.0263
D12S391	24	120	0.8654	0.0279
FGA	29	111	0.8702	0.0299
Penta D*	16	70	0.8733	0.0360
D21S11	32	98	0.8331	0.0399
D19S433	16	83	0.8100	0.0534
D8S1179	11	48	0.7966	0.0553
vWA	11	42	0.8000	0.0624
D16S539	9	30	0.7812	0.0723
D13S317	9	30	0.7749	0.0724
D7S820	12	35	0.7826	0.0745
TH01	9	27	0.7518	0.0752
D2S441	14	46	0.7777	0.0807
D10S1248	12	41	0.7812	0.0828
D3S1358	11	31	0.7489	0.0904
D22S1045	11	45	0.7567	0.0935
D5S818	9	34	0.7225	0.1057
CSF1PO	10	33	0.7567	0.1071
TPOX	10	30	0.6830	0.1351

Rank ordered
by their variability
(P_I = probability of identity)

Better for mixtures (more alleles seen)

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

Better for kinship (low mutation rate)

Concordance Studies

Commercially Available STR Kits

Applied Biosystems (17)

- ~~AmpFlSTR Blue (1996)~~
- ~~AmpFlSTR Green I (1997)~~
- Profiler (1997)
- Profiler Plus (1997)
- COfiler (1998)
- SGM Plus (1999)
- **Identifiler** (2001)
- Profiler Plus ID (2001)
- ~~SEfiler (2002)~~
- Yfiler (2004)
- MiniFiler (2007)
- SEfiler Plus (2007)
- Sinofiler (2008) – China only
- **Identifiler Direct** (2009)
- NGM (2009)
- **Identifiler Plus** (2010)
- NGM SElect (2010)

Promega Corporation (13)

- PowerPlex 1.1 (1997)
- PowerPlex 1.2 (1998)
- PowerPlex 2.1 (1999)
- **PowerPlex 16** (2000)
- PowerPlex ES (2002)
- PowerPlex Y (2003)
- PowerPlex S5 (2007)
- **PowerPlex 16 HS** (2009)
- PowerPlex ESX 16 (2009)
- PowerPlex ESX 17 (2009)
- PowerPlex ESI 16 (2009)
- PowerPlex ESI 17 (2009)
- PowerPlex 18D (2010)

Qiagen (2010)

Primarily selling kits in Europe

Due to patent restrictions cannot sell in U.S.

- ESSplex
- ESSplex SE
- Decaplex SE
- IDplex
- Nonaplex ESS
- Hexaplex ESS
- HD (Chimera)
- Argus X-12
- Argus Y-12
- DIPlex (30 indels)

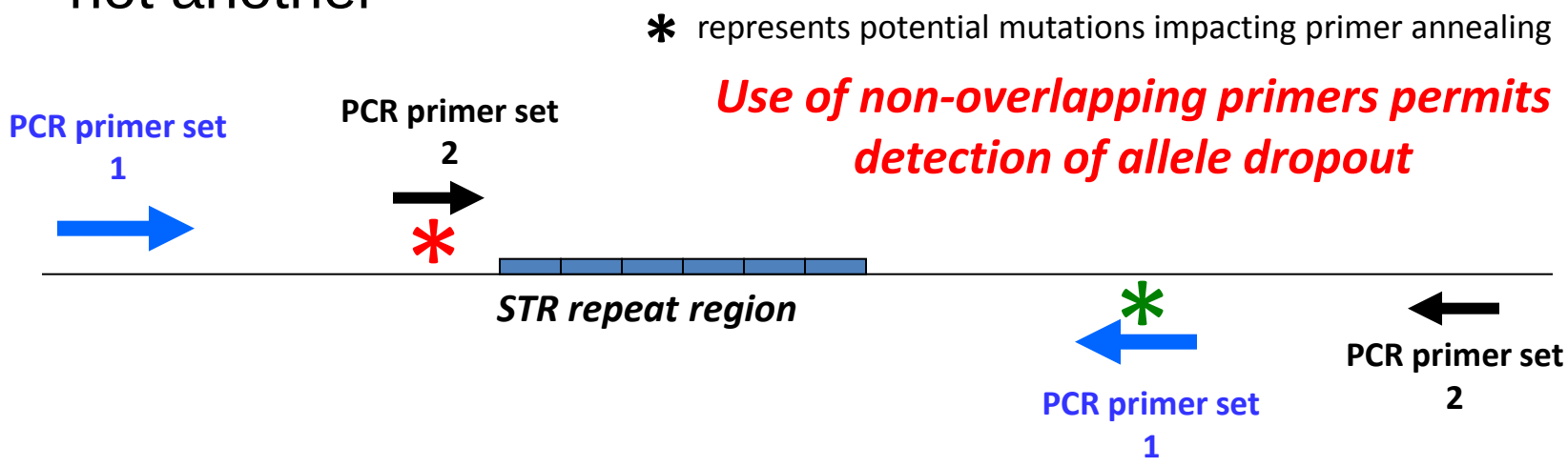
~1/3 of all STR kits were released in the last year

STR Kit Concordance Testing

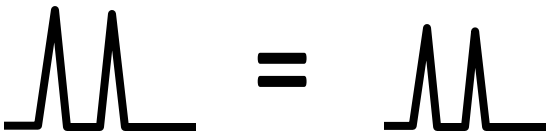
- Many of these STR kits have different primer sequences for amplifying the same STR locus
- Need to analyze the same DNA samples with different STR typing kits looking for differences
- In some rare cases, allele dropout (null alleles) may occur due to mutations in primer binding regions

Purpose of Concordance Studies

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



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

STR Kit Concordance Testing

Profiles in DNA Article Published April 2010

Article Type: Feature

Volume 13 No. 1, April 2010

Strategies for Concordance Testing

Carolyn R. Hill, Margaret C. Kline, David L. Duewer and John M. Butler
National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, Maryland, USA

Concordance evaluations are important to conduct to determine if there are any allelic dropout or "null alleles" present in a data set. These studies are performed because there are a variety of commercial short tandem repeat (STR) multiplex kits with different configurations of STR markers available to the forensic community. The placement of the markers can vary between kits because the primer sequences were designed to amplify different polymerase chain reaction (PCR) product sizes. When multiple primer sets are used, there is concern that allele dropout may occur due to primer-binding-site mutations that affect one set of primers but not another.

http://www.promega.com/profiles/1301/1301_08.html

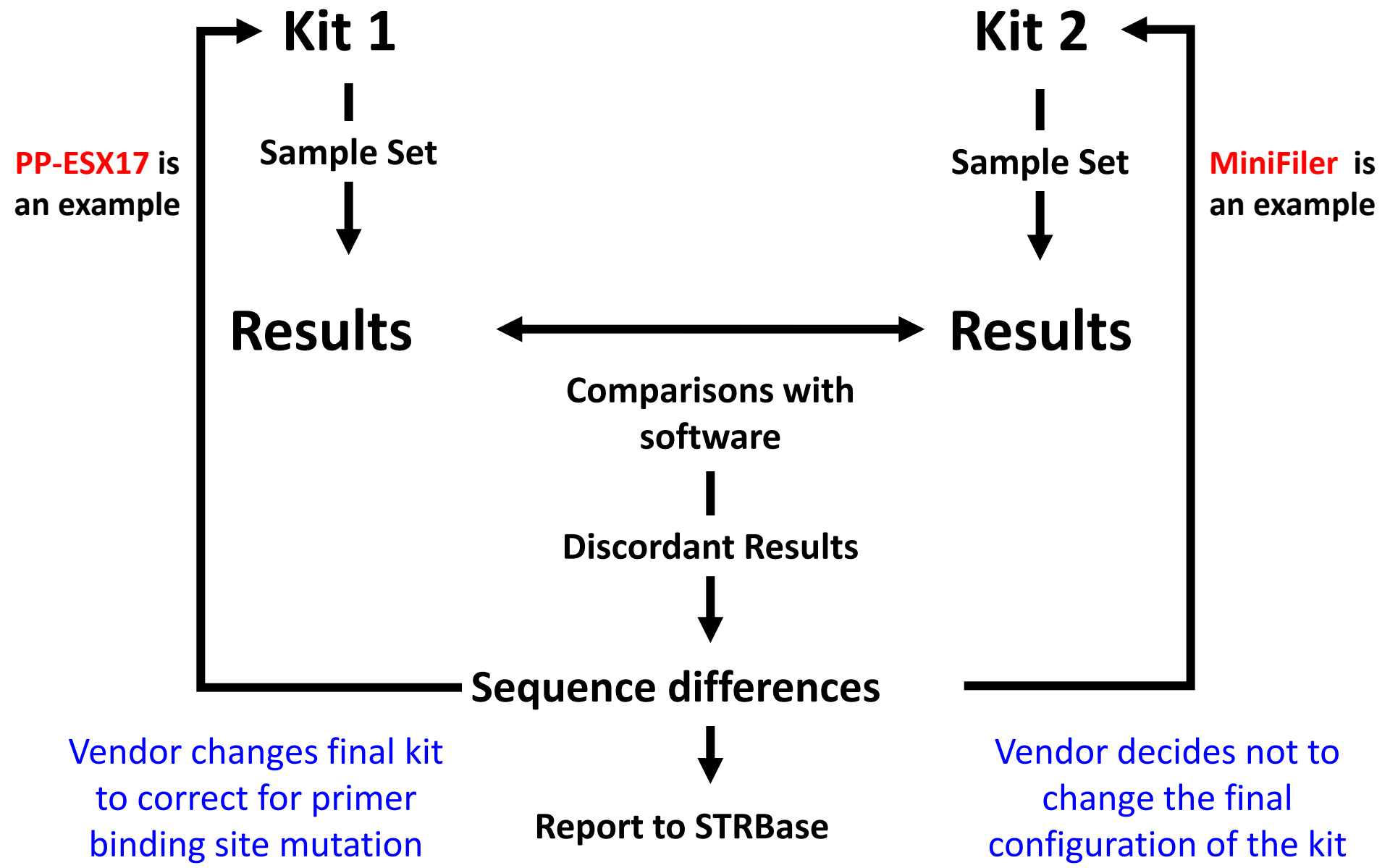
The 4 “S’s” of Concordance

- NIST Standard Samples
  - Run same samples with multiple kits to compare results
- Concordance Software
  - Allows comparison of data sets using NIST developed software

    http://www.cstl.nist.gov/biotech/strbase/software.htm
- DNA Sequencing
  - To validate and determine the exact cause for the null allele
- STRBase website
  - To report verified null alleles and discordant results to the forensic community

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

NIST Concordance Testing Steps



NIST Sample Set (>1450 Samples)

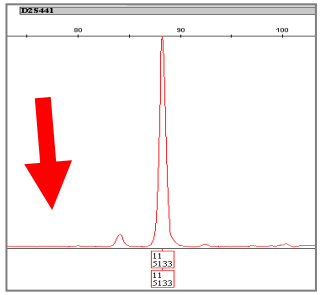
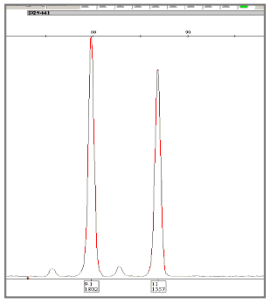
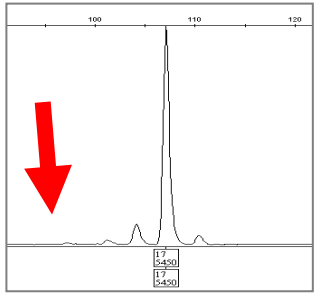
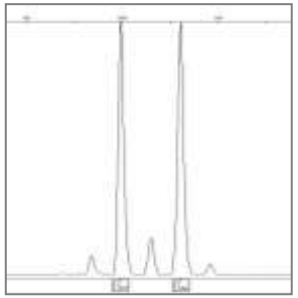
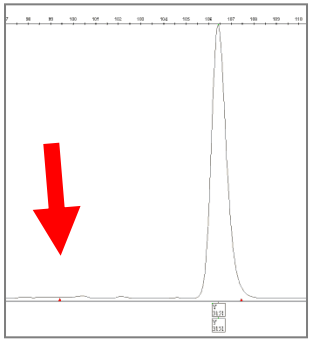
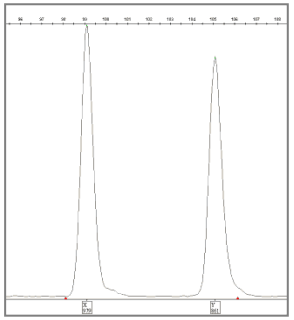
- **NIST U.S. population samples**
  - 260 African American, 260 Caucasian, 140 Hispanic, 3 Asian

- **U.S. father/son paired samples**
  - ~**100 fathers/100 sons for each group**: 200 African American, 200 Caucasian, 200 Hispanic, 200 Asian

- **NIST SRM 2391b**, PCR-based DNA Profiling Standard (highly characterized)
  - 10 genomic DNA samples, 2 cell line samples
  - Includes 9947A and 9948

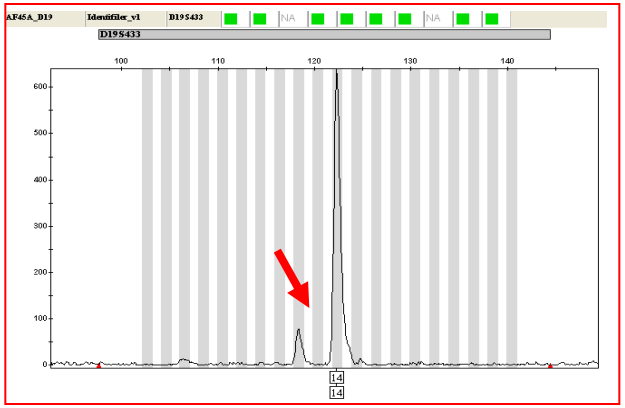
- **NIST SRM 2391c**, PCR-based DNA Profiling Standard
  - 4 genomic DNA (one mixture)
  - 2 cell lines (903 and FTA paper)

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**

D19S433 Discordance

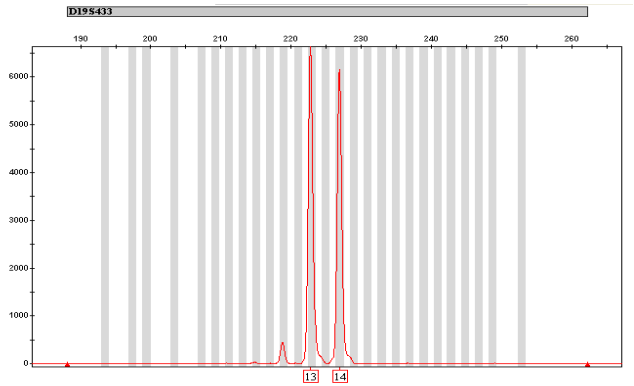
Identifiler & NGM = 14,14



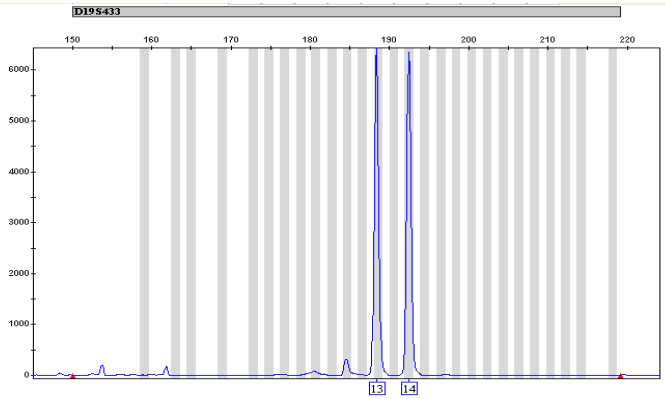
AF45A (Asian)

Allele 13 was missing in **two different Asian samples** with ABI primers
= 2/2886 =
0.07% discordance

ESX 17 = 13,14



ESI 17 = 13,14



Frequencies [for] the silent allele were determined to be 0.0114 in 176 people from Shizuoka (Honshu) and 0.0128 in 156 people from Okinawa

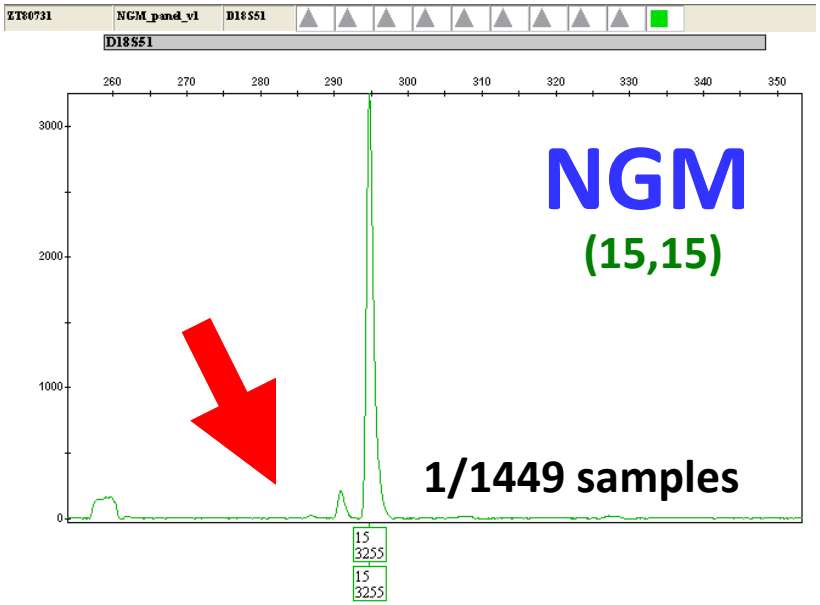
J Forensic Sci, September 2008, Vol. 53, No. 5
doi: 10.1111/j.1556-4029.2008.00806.x
Available online at: www.blackwell-synergy.com

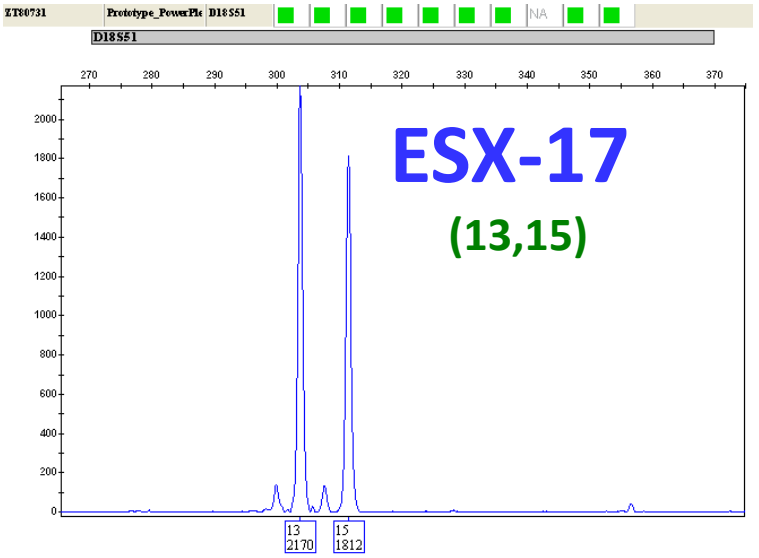
Natsuko Mizuno,[1] D.V.M.; Tetsushi Kitayama,[1] M.Sc.; Koji Fujii,[1] Ph.D.; Hiroaki Nakahara,[1] D.V.M.; Kanako Yoshida,[1] Ph.D.; Kazumasa Sekiguchi,[1] Ph.D.; Naoto Yonezawa,[2] Ph.D.; Minoru Nakano,[2] Ph.D.; and Kentaro Kasai,[1] Ph.D.

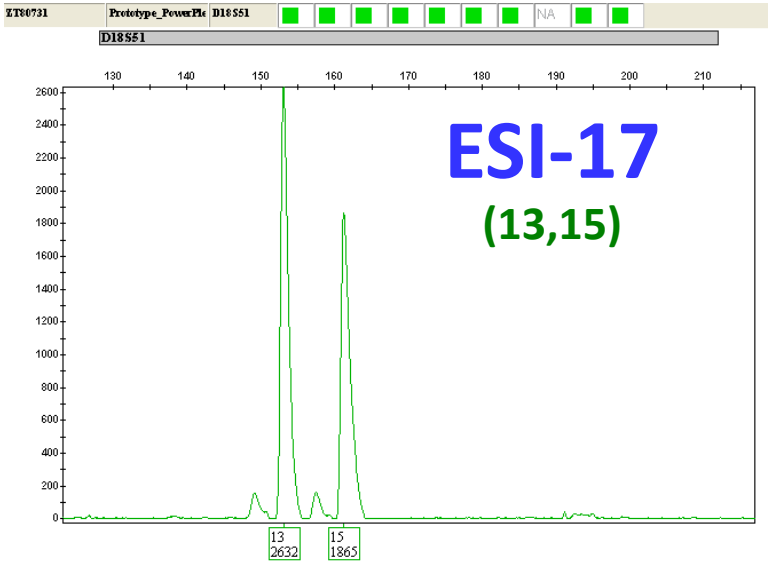
A D19S433 Primer Binding Site Mutation and the Frequency in Japanese of the Silent Allele It Causes

T→A SNP 8 bp downstream impacting reverse primer binding with Identifiler (and thus SGM Plus)

D18S51 Null Allele



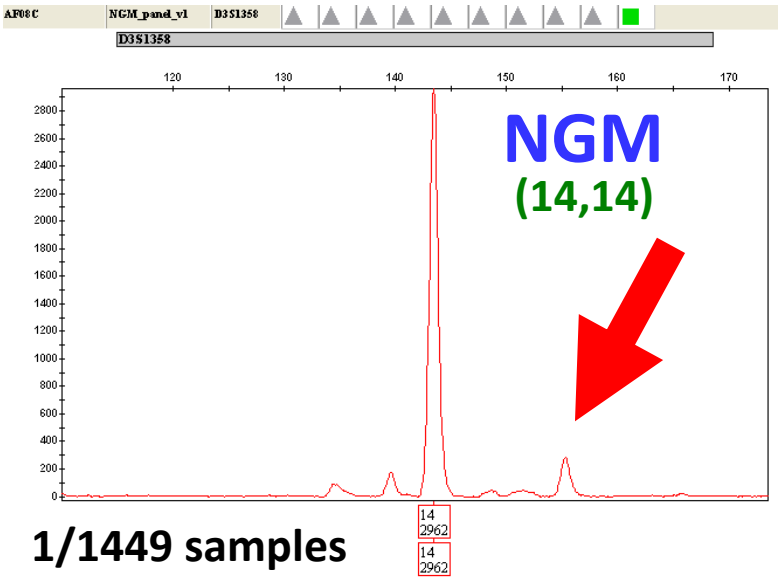




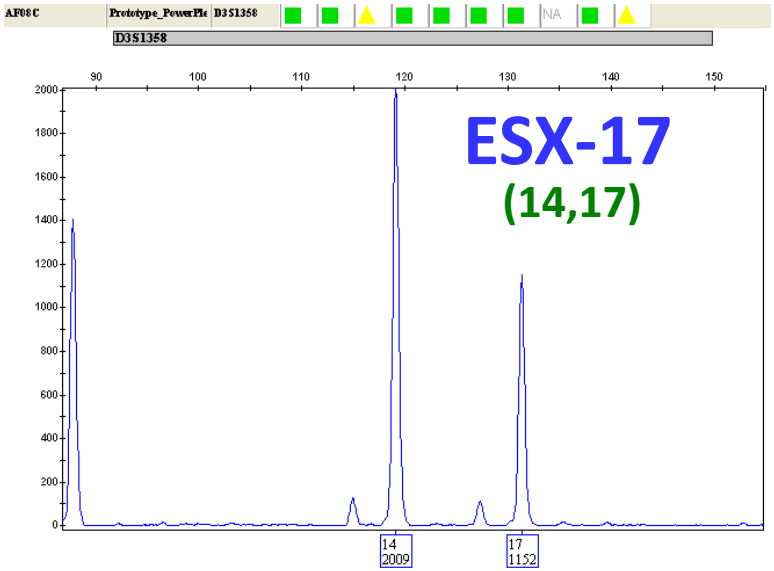
Correct type (13,15)

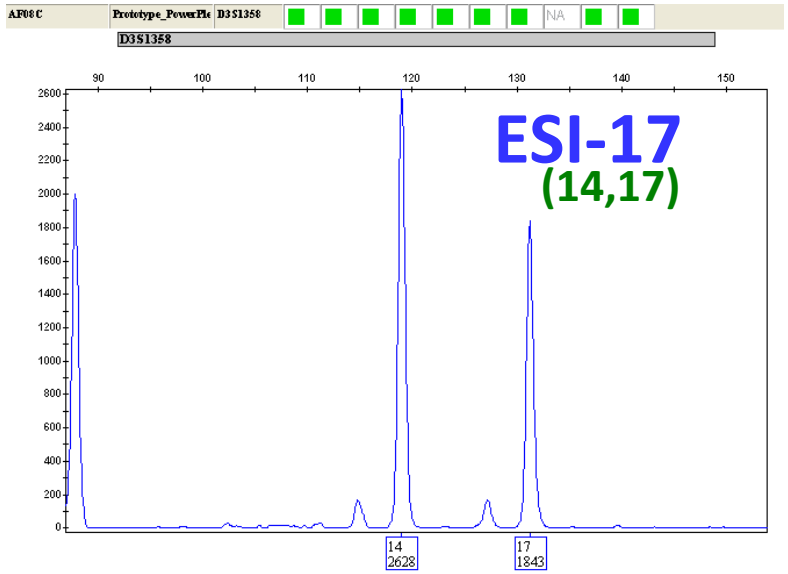
C→T SNP 172 bp downstream from repeat

D3S1358 Null Allele



1/1449 samples





Correct type (14,17)

G → C SNP 11 bp downstream from repeat

Kits compared	Samples	Loci Compared	Comparisons	# Differences	Concordance (%)
ID-SGM+	1424	11	15,664	1	99.994
ID-Pro+	1415	10	14,150	1	99.993
ID-IDplex	1426	16	22,816	29	99.873
ID-PP16	662	14	9,268	4	99.957
ID-MiniFiler	1137	9	10,233	26	99.746
ID-NGM	1437	11	15,807	3	99.981
ID-NGMs	663	11	7,293	0	100.000
ID-ESX17	1443	11	15,873	5	99.968
ID-ESI17	1443	11	15,873	4	99.975
ID-ESSplex	1433	11	15,763	28	99.822
ID-ESSplexSE	662	11	7,282	17	99.767
ID-Hexaplex	653	2	1,306	1	99.923
PP16-SGM+	651	9	5,859	1	99.983
PP16-Pro+	647	10	6,470	2	99.969
PP16-IDplex	657	14	9,198	3	99.967
PP16-MiniFiler	656	8	5,248	14	99.733
PP16-NGM	657	9	5,913	3	99.949
PP16-NGMs	662	9	5,958	1	99.983
PP16-ESX17	662	9	5,958	1	99.983
PP16-ESI17	662	9	5,958	0	100.000
PP16-ESSplex	653	9	5,877	16	99.728
PP16-ESSplexSE	662	9	5,958	16	99.731
PP16-Hexaplex	653	2	1,306	1	99.923
SGM+-Pro+	1415	7	9,905	0	100.000
SGM+-IDplex	1424	11	15,664	5	99.968
SGM+-MiniFiler	1137	6	6,822	10	99.853
SGM+-NGM	1424	11	15,664	4	99.974
SGM+-NGMs	651	11	7,161	0	100.000
SGM+-ESX17	1424	11	15,664	6	99.962
SGM+-ESI17	1424	11	15,664	5	99.968
SGM+-ESS	1424	11	15,664	5	99.968
SGM+-ESSplexSE	651	11	7,161	5	99.930
SGM+-Hexaplex	651	2	1,302	1	99.923
Pro+-IDplex	1415	10	14,150	5	99.965
Pro+-MiniFiler	1137	6	6,822	16	99.765
Pro+-NGM	1415	7	9,905	4	99.960
Pro+-NGMs	647	7	4,529	0	100.000
Pro+-ESX17	1415	7	9,905	4	99.960
Pro+-ESI17	1415	7	9,905	3	99.970
Pro+-ESS	1415	7	9,905	4	99.960
Pro+-ESSplexSE	647	7	4,529	4	99.912
Pro+-Hexaplex	647	1	647	1	99.845
IDplex-MiniFiler	1137	9	10,233	48	99.531
IDplex-NGM	1426	11	15,686	30	99.809
IDplex-NGMs	657	11	7,227	17	99.765
IDplex-ESX17	1426	11	15,686	28	99.821
IDplex-ESI17	1426	11	15,686	27	99.828
IDplex-ESS	1426	11	15,686	1	99.994
IDplex-ESSplexSE	657	11	7,227	1	99.986
IDplex-Hexaplex	653	2	1,306	1	99.923
MiniFiler-NGM	1137	6	6,822	13	99.809
MiniFiler-NGMs	656	6	3,936	10	99.746
MiniFiler-ESX17	1137	6	6,822	10	99.853
MiniFiler-ESI17	1137	6	6,822	9	99.868
MiniFiler-ESS	1137	6	6,822	35	99.487
MiniFiler-ESSplexSE	656	6	3,936	35	99.111
MiniFiler-Hexaplex	653	1	653	1	99.847
NGM-NGMs	657	16	10,512	14	99.867
NGM-ESX17	1437	16	22,992	16	99.930
NGM-ESI17	1437	16	22,992	18	99.922
NGM-ESS	1433	16	22,928	42	99.817
NGM-ESSplexSE	657	16	10,512	22	99.791
NGM-Hexaplex	653	7	4,571	9	99.803
NGMs-ESX17	662	17	11,254	4	99.964
NGMs-ESI17	662	17	11,254	14	99.876
NGMs-ESS	653	16	10,448	17	99.837
NGMs-ESSplexSE	662	17	11,254	34	99.698
NGMs_Hexaplex	653	7	4,571	3	99.934
ESX17-ESI17	1443	17	24,531	19	99.923
ESX17-ESS	653	16	10,448	34	99.675
ESX17-ESSplexSE	662	17	11,254	25	99.778
ESX17-Hexaplex	657	7	4,599	6	99.870
ESI17-ESS	653	16	10,448	28	99.732
ESI17-ESSplexSE	662	17	11,254	30	99.733
ESI17-Hexaplex	657	7	4,599	3	99.935
ESS-ESSplexSE	653	16	10,448	0	100.000
ESS-Hexaplex	653	7	4,571	3	99.934
ESSplexSE-Hexaplex	653	7	4,571	3	99.934
SE33-ESX17	1443	1	1,443	6	99.584
SE33-ESI17	1443	1	1,443	17	98.822
SE33-NGMs	663	1	663	4	99.397
SE33-ESSplexSE	662	1	662	21	96.828
ESI17p-ESX17	477	17	8,109	7	99.914
ESI17p-NGMs	477	17	8,109	2	99.975
ESI17p-ESSplexSE	477	17	8,109	42	99.482
ESI17p-SE33	477	1	477	4	99.161
PP18D-ID	50	16	800	2	99.750
PP18D-PP16	703	16	11,248	1	99.991
ESX17*/ESX17	1443	17	24531	4	99.984
ESX17*/ESI17p	477	17	8109	3	99.963
ESX17*/NGM	1437	16	22992	22	99.904
ESX17*/NGMs	663	17	11271	4	99.965
ESX17*/ESS	1433	16	22928	30	99.869
ESX17*/ESSplexSE	662	17	11254	44	99.609
ESX17*/Hexaplex	653	7	4571	2	99.956
26plex/ESX17	1443	3	4329	4	99.908
26plex/ESI17	1443	3	4329	0	100.000
26plex/NGM	1437	3	4311	11	99.745
26plex/NGMs	663	3	1989	0	100.000
26plex/ESS	1433	3	4299	0	100.000
26plex/ESSplexSE	662	3	1986	0	100.000
26plex/Hexaplex	653	3	1959	2	99.898
26plex/ESX17*	663	3	1989	0	100.000
miniSTRs/ESX17	663	3	1989	3	99.849
miniSTRs/ESI17	663	3	1989	0	100.000
miniSTRs/NGM	657	3	1971	3	99.848
miniSTRs/NGMs	663	3	1989	0	100.000
miniSTRs/ESS	653	3	1959	0	100.000
miniSTRs/ESSplexSE	662	3	1986	0	100.000
miniSTRs/Hexaplex	653	3	1959	2	99.898
miniSTRs/ESX17*	663	3	1989	0	100.000
Totals	102345	1021	948301	1109	**99.883**

Completed Concordance Studies

Kits compared	Samples	Loci Compared	Comparisons	# Differences	Concordance (%)
111	102,345	1,021	948,301	1,109	99.883

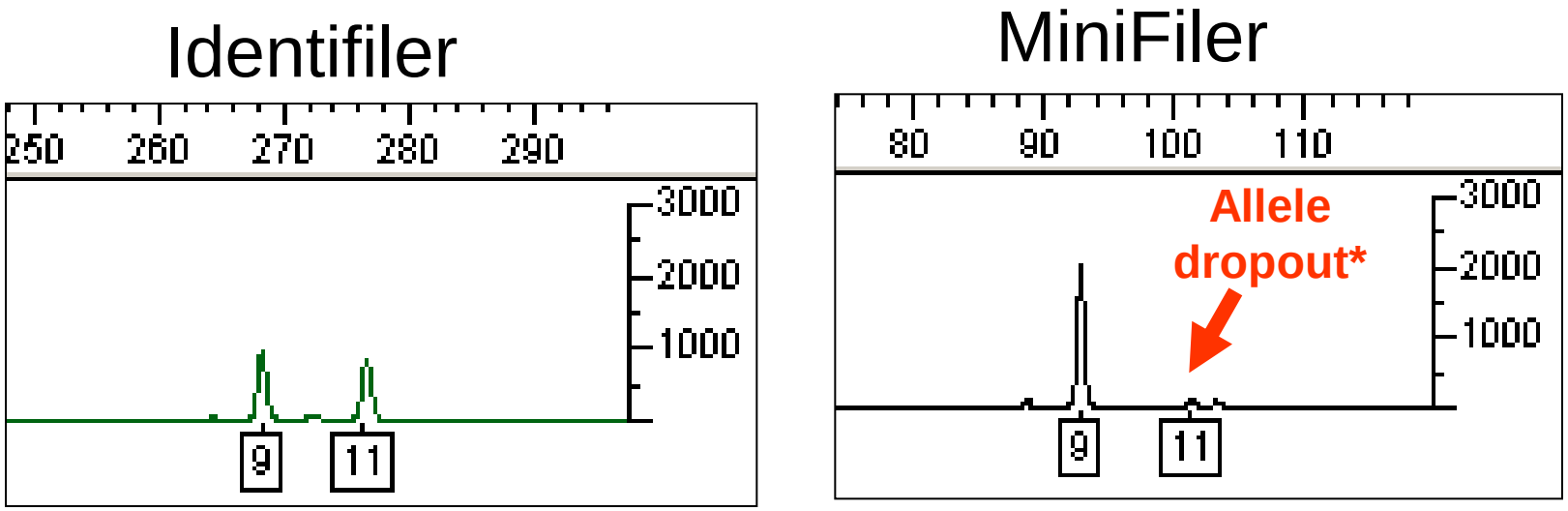
948,301 allele comparisons

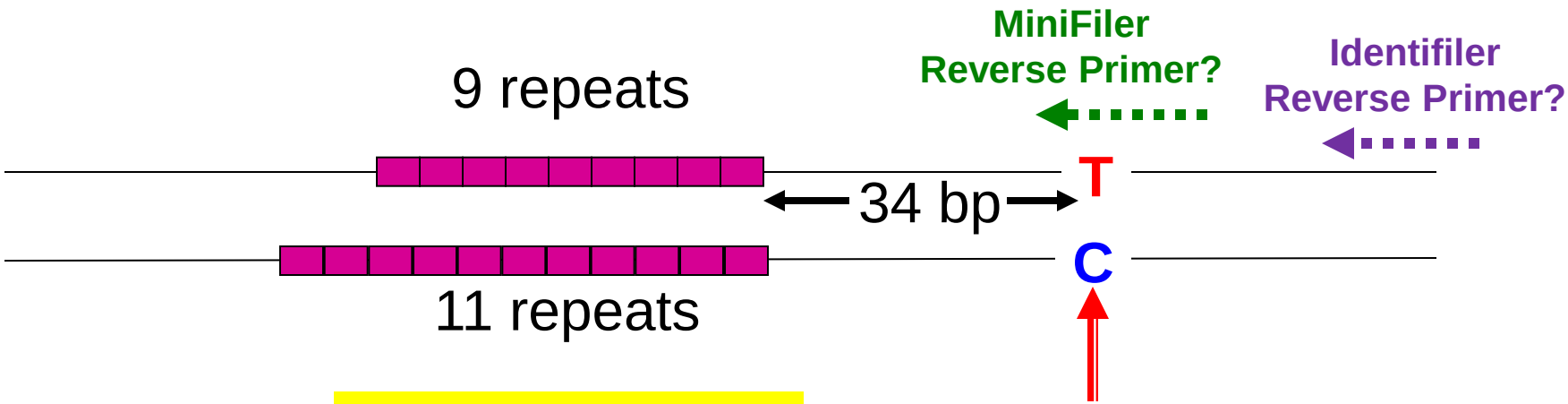
1,109 total differences

99.88% concordance

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

SRM 2391b Genomic 8 with D16S539

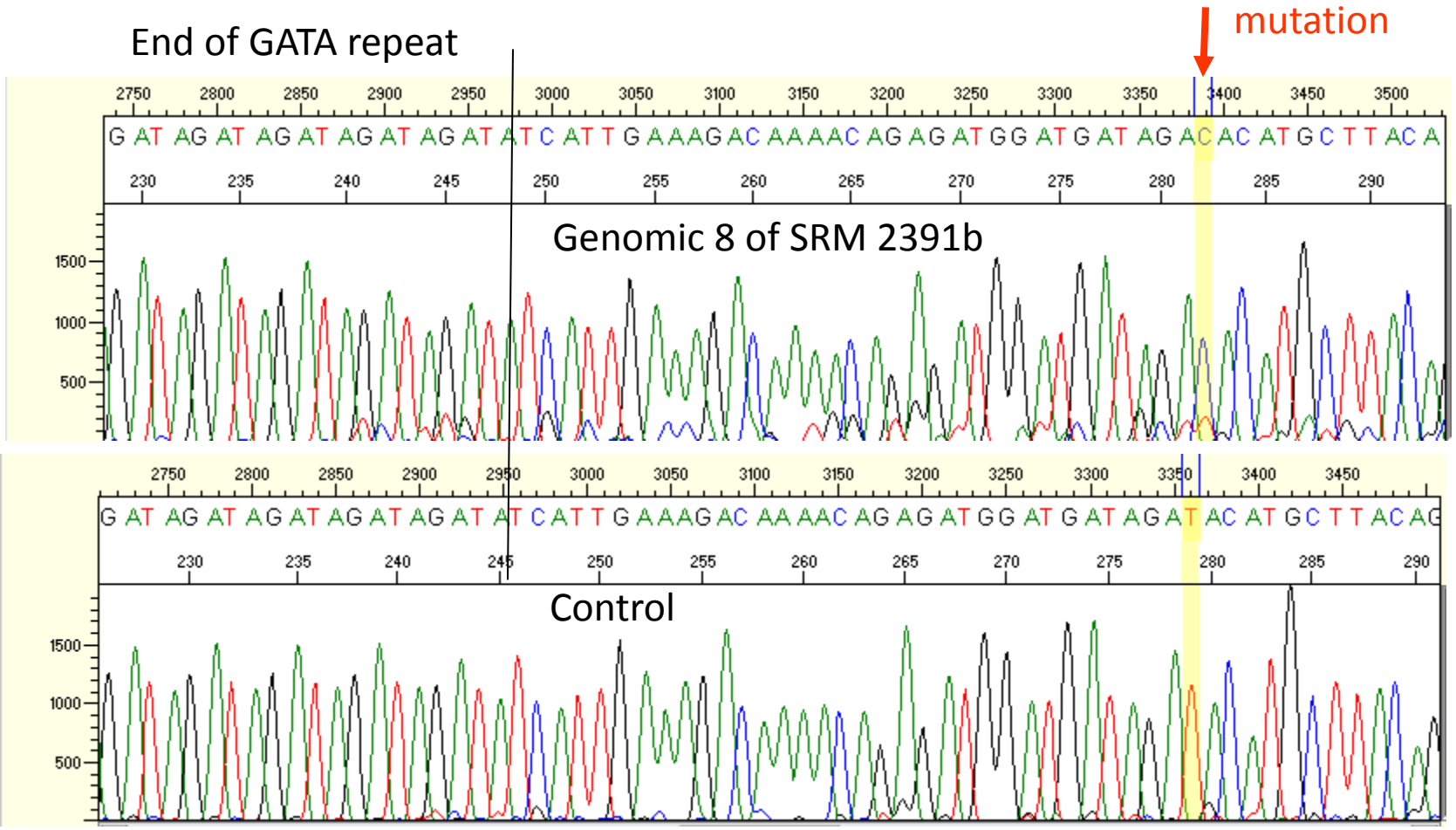




Type 9T,11C

D16S539 SRM 2391b Genomic 8

T→C mutation 34 bp downstream of the repeat



Position of the T→C probably affects the reverse primer of Minifiler and is the 3rd base found the 5'end of the Reverse PP16 primer. This could explain the imbalance of the allele seen when using PP16.

miniSTRs and the 26plex

More Loci are Useful
in Situations Involving Relatives

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

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

Selection of New Autosomal Loci

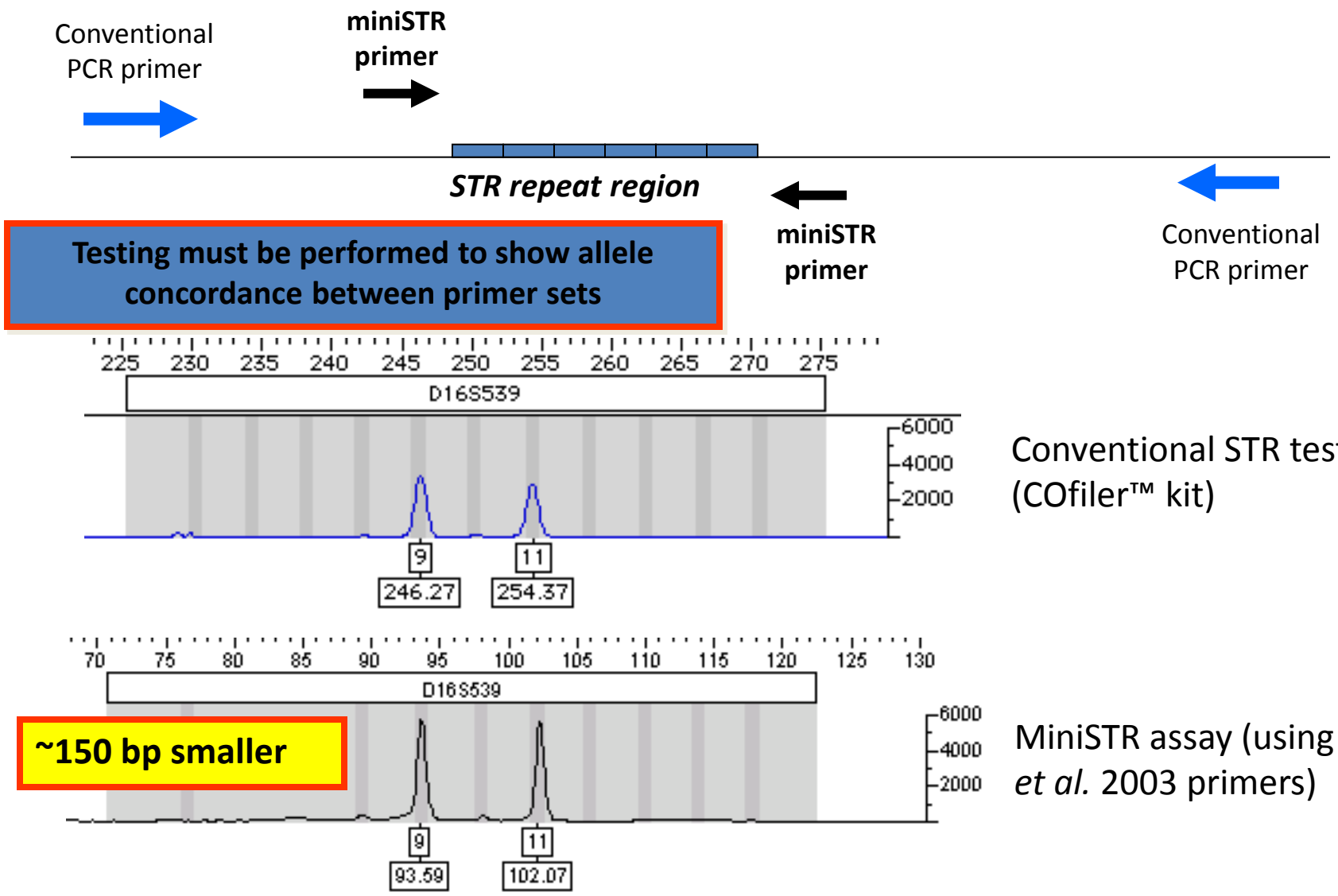
- Aim to have candidate sets for optimal **miniSTRs**
  - Obtaining additional information with degraded DNA samples

- Using ~900 STR loci with some literature data as a starting point…
  - Loci with high heterozygosities (>0.7)
  - Loci with small allele ranges (<24 bp) – low mutation?
  - Tetra (some tri-)nucleotide repeats without variants
  - Clean flanking regions (PCR products <140 bp)

- **26 loci** met criteria and were fully characterized…

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

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



Characterizing New Loci

- New loci were chosen based on the following characteristics:
  - Genomic Position
  - Polymorphic Content
  - Span/Range of observed alleles
- Details about the characterization process have all been previously reported at length:

Our publications and presentations are made available at:

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

  - *Eighteenth International Symposium on Human Identification*. See http://www.promega.com/geneticidproc/
  - John Butler's talk at the 18th International Symposium on Human Identification (Promega 2007), "New Autosomal and Y-Chromosome STR Loci: Characterization and Potential Uses"

Characterization of New Loci

"Computer Work"

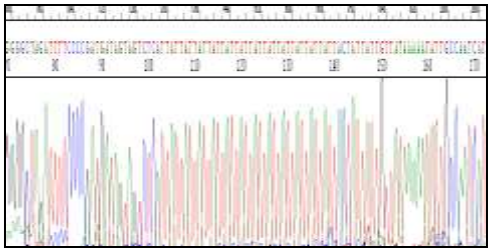
Candidate STR marker selection → **Pull down sequence data from the web** → **Identify Chromosome Location** → **Screen for PCR Primers** → **Test primers for Multiplex-ability**

(e.g. literature searches) (e.g. NCBI) (e.g. Human BLAT Search) (e.g. Primer3) (e.g. AutoDimer - NIST)

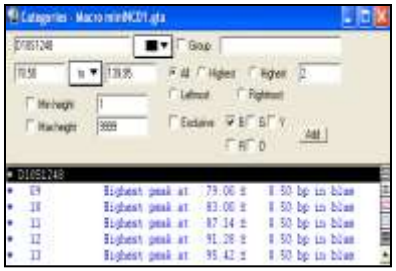
"Laboratory Work"

Test Markers on Population samples

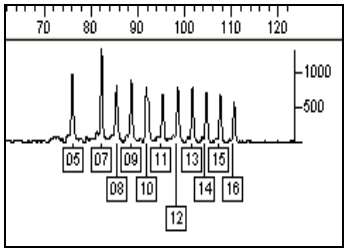
Sequence homozygotes to determine allele sizes



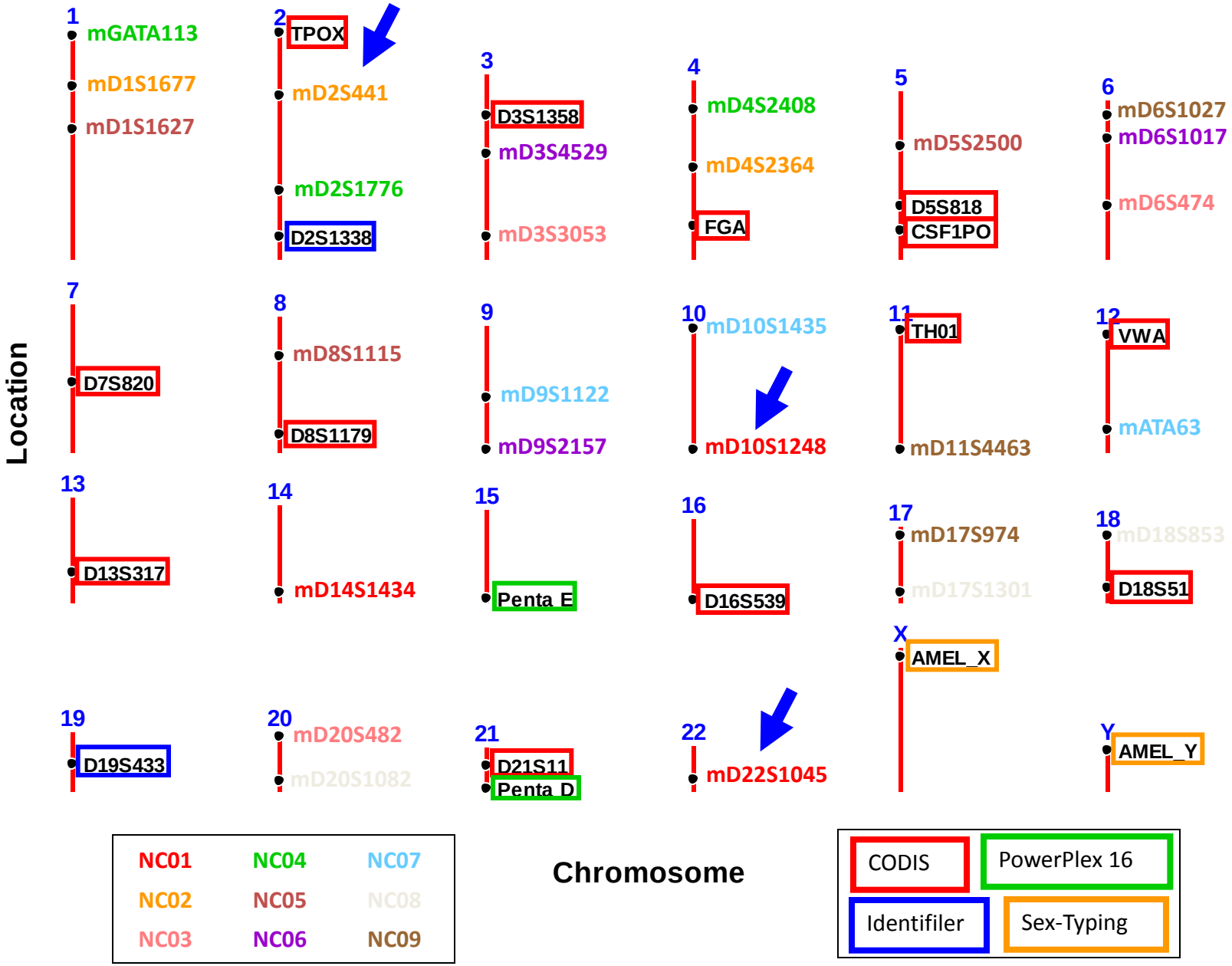
Build Macros for Genotyping



Construct Allelic Ladders



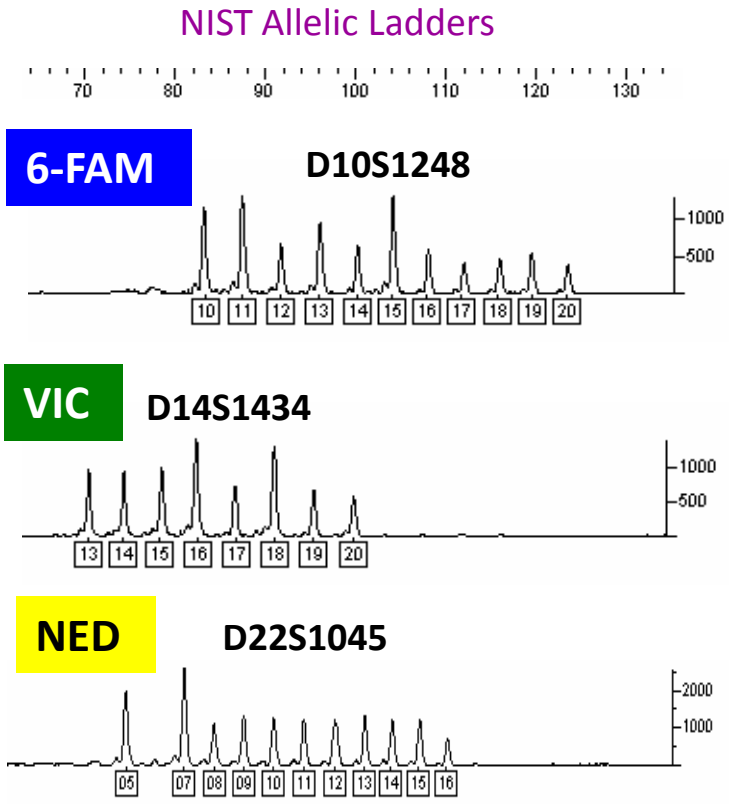
Chromosomal Locations of New miniSTR Loci



Multiple Miniplexes

- **26 characterized loci** divided into 10 miniplexes
- One locus per dye color
- Allelic ladders created
- **Amplicons <140 bp**
- miniSTRs
- Work with 100 pg DNA
- **For degraded samples**
  (bones in missing persons cases)

NC = Non-CODIS or non-core



NC01 Loci

See Dixon et al. (2006) *Forensic Sci. Int.* 164: 33-44.

NC Miniplexes

NC01
D10S1248
D14S1434
D22S1045

NC02
D1S1677
D2S441

NC03
D3S3053
D6S474

26 total new loci

NC04
D1GATA113
D2S1776
D4S2408

NC05
D1S1627
D8S1115
D9S324

NC06
D3S4529
D9S2157
D10S1430

NC10
D3S3053
D6S474
D20S482

NC07
D9S1112
D12ATA63
D14S1280

NC08
D17S1301
D18S8534
D20S1082

NC09
D10S2327
D11S4463
D17S974

Removed because they were problematic

In Jan 2008 Issue of *J. Forensic Sci.*

J. Forensic Sci., Jan 2008, 53(1):73-80

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

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

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

- Characterization of **26** new autosomal loci
- Primer sequences, GeneMapper bins and panels, genotypes on common samples, and allele frequency information **already available on STRBase**

http://www.cstl.nist.gov/div831/strbase/miniSTR.htm
http://www.cstl.nist.gov/div831/strbase/newSTRs.htm

European Labs Have Adopted the NIST-Developed NC (non-CODIS) miniSTRs

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

Short communication

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

These 3 loci are included in the new European multiplex kits: Applied Biosystems NGM kits and the Promega PowerPlex ESX 16/17 and ESI 16/17 systems

[c] *Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Denmark*
[d] *Institute of Legal Medicine, University of Cologne, Germany*

Received 25 May 2005; accepted 26 May 2005

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

(D14 has been replaced with D2S441 from NC02)

The Design of the Multiplex

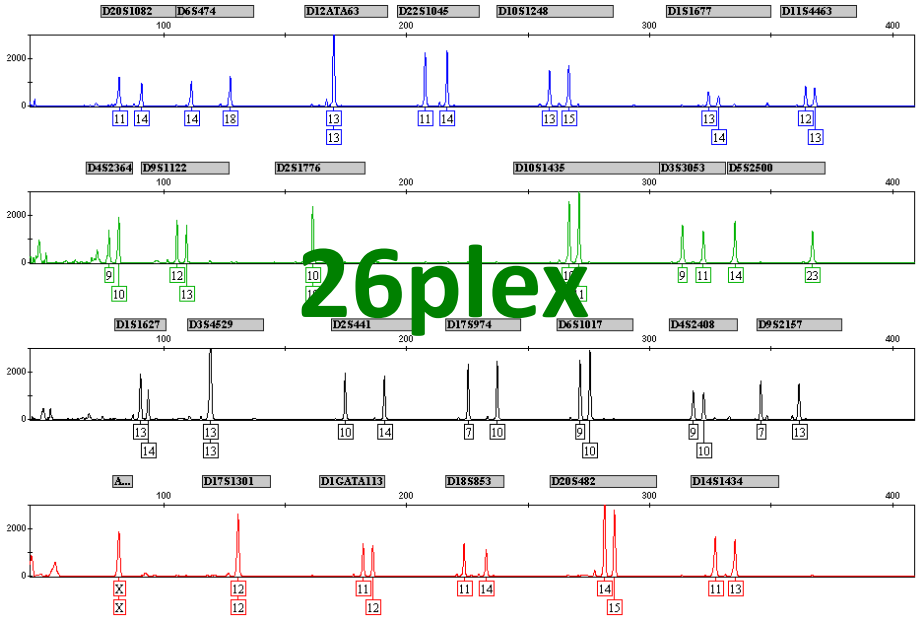
- **Goal**: A single amplification 5-dye multiplex to combine the 26 new autosomal loci + Amelogenin in one reaction (27plex)
- How can this be achieved?
  - Initial placement of all loci within 6FAM, VIC, NED, and PET dye channels (the size standard is in the LIZ channel)
  - Primer redesign for all but 7 of the original miniSTR loci
  - Trial and error of primer compatibility, as well as balancing for all working primers

26plex STR Multiplex

- So far **25 STRs and amelogenin** in single multiplex (Eventual goal to have all 26 loci)
- Multiple loci in four dye channels
- **Amplicons 70 to 400 bp**

  (No longer ‘miniSTRs’)
- Typically use 1 ng DNA, 30 cycles
- **For reference samples**
  (a missing person’s relatives)

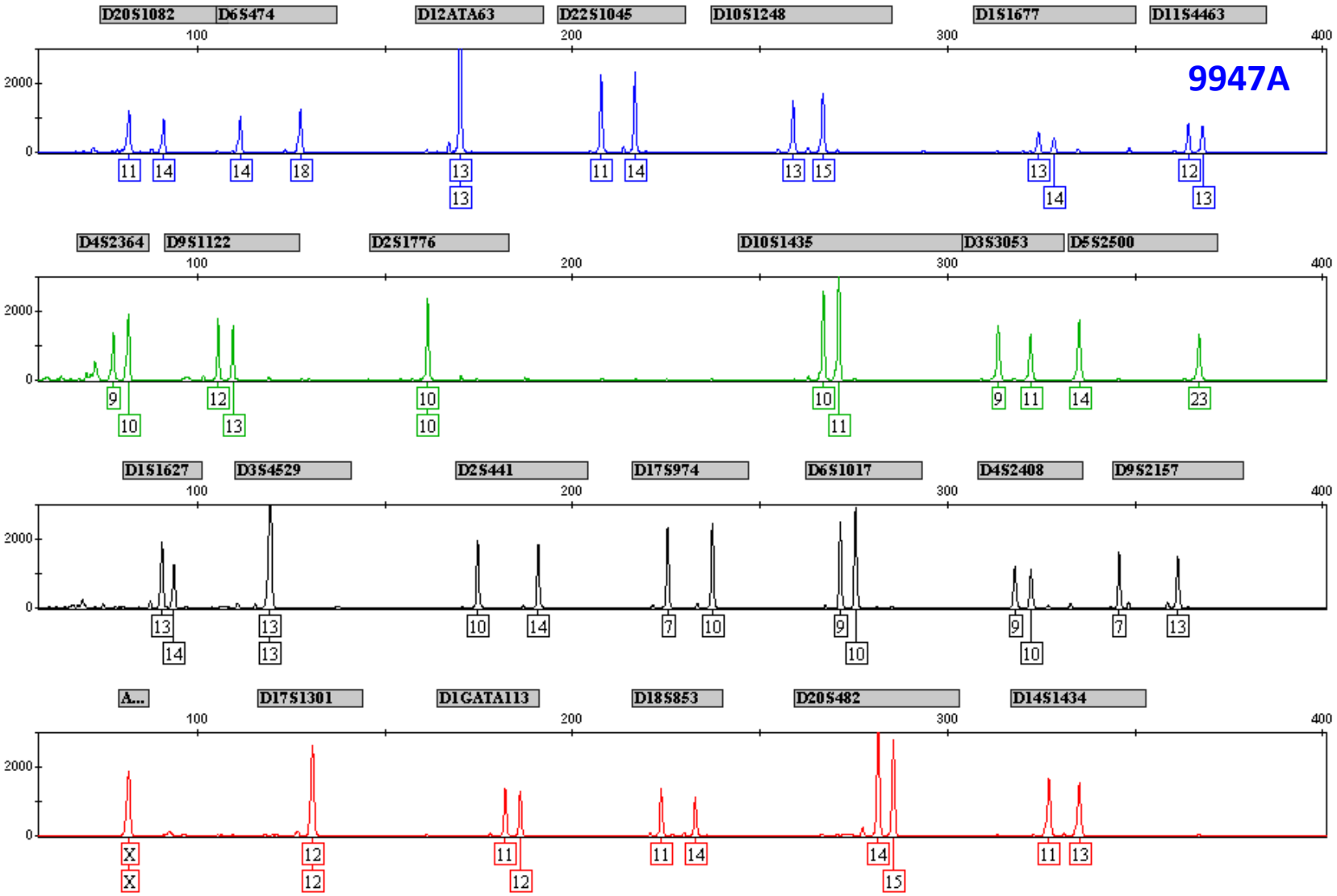
“Autoplex” or
“miniMegaplex”



All loci unlinked from core (CODIS) STRs

NIST STR 26plex

Hill et al. (2009) *Journal of Forensic Sciences*



Gender identification + 25 autosomal STR loci in a single amplification

In Sept 2009 Issue of *J. Forensic Sci.*

J. Forensic Sci., Sept 2009, 54(5):1008-1015

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*†

- Strategies for building multiplexes
- Primer sequences and PCR conditions listed
- GeneMapper bins and panels, genotypes on common samples, and allele frequency information **available on STRBase**

http://www.cstl.nist.gov/biotech/strbase/str26plex.htm
http://www.cstl.nist.gov/div831/strbase/newSTRs.htm

Low Template DNA (LT-DNA)

Some Definitions of Low Template (LT) DNA

- Working with **<100-200 pg genomic DNA**
- Considered to be data below stochastic threshold level where PCR amplification is not as reliable (determined by each laboratory; typically 150-250 RFUs)
- Enhancing the sensitivity of detection (increasing PCR cycles, PCR product clean-up, increasing CE injection/voltage)
- Having too few copies of DNA template to ensure reliable PCR amplification (allelic or full locus drop-out)
- Can often be the minor component of mixture samples consisting of low level DNA template amounts

Impact of DNA Amount into Multiplex PCR Reaction

We generally aim for 0.5-2 ng

DNA amount
(log scale)

100 ng

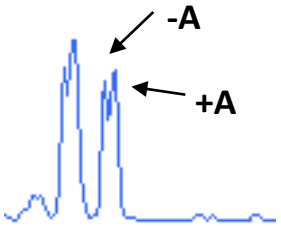
10 ng

1 ng

0.1 ng

0.01 ng

High levels of DNA create interpretation challenges (more artifacts to review)



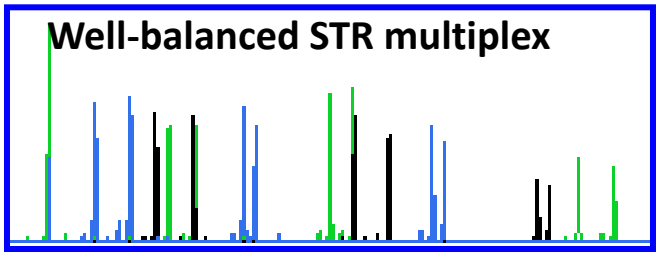
Too much DNA

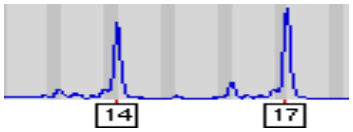
- Off-scale peaks
- Split peaks (+/-A)
- Locus-to-locus imbalance

2.0 ng

STR Kits Work Best in This Range

0.5 ng

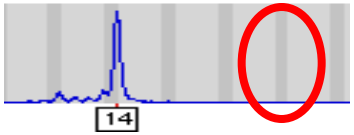




100 pg template

Too little DNA

- Heterozygote peak imbalance
- Allele drop-out
- Locus-to-locus imbalance



5 pg template

Stochastic effects when amplifying low levels of DNA can produce allele dropout

Challenges of LT-DNA Testing

Gill, P. (2001) *Croatian Med. J.* 42(3): 229-232

- Increased chance for contamination (want a sterile lab environment to reduce staff contamination)

- Data interpretation is more complicated (due to stochastic variation during PCR amplification):
  - Heterozygote peak imbalance
  - Allele drop-out
  - Allele drop-in
  - Increased stutter products

LT-DNA profiles should be interpreted with careful guidelines

Signal Enhancement Techniques

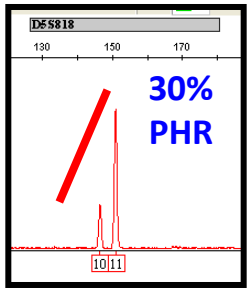
- Additional PCR cycles
- More sensitive kits (Identifiler Plus and PowerPlex 16 HS)
- Microcon cleanup to remove salts that interfere with electrokinetic injection (MinElute PCR Purification Kit from Qiagen)
- Lower PCR volume (concentrates amplicon)
- Increase TaqGold/enzyme concentration
- Longer CE injection times and voltage
  - 10 s @ 3 kV = 30
  - 5 s @ 2 kV = 10

Stochastic (Random) Effects with LT-DNA

When Combined with Higher Sensitivity Techniques

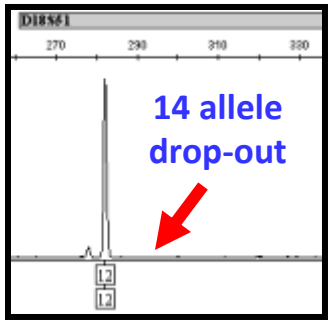
Loss of True Signal (***False Negative***)

Heterozygote Peak Imbalance



Identifiler, 30 pg DNA, 31 cycles

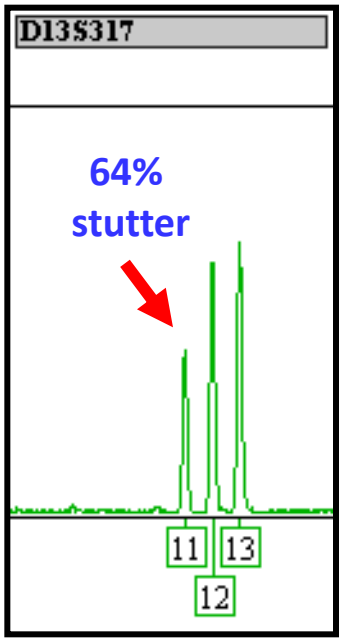
Allelic **Drop-out**



Identifiler, 30 pg DNA, 31 cycles

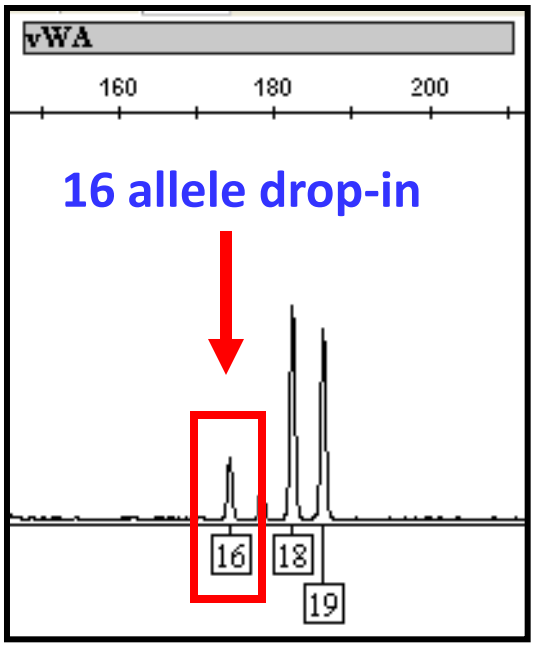
Gain of False Signal (***False Positive***)

Higher Stutter



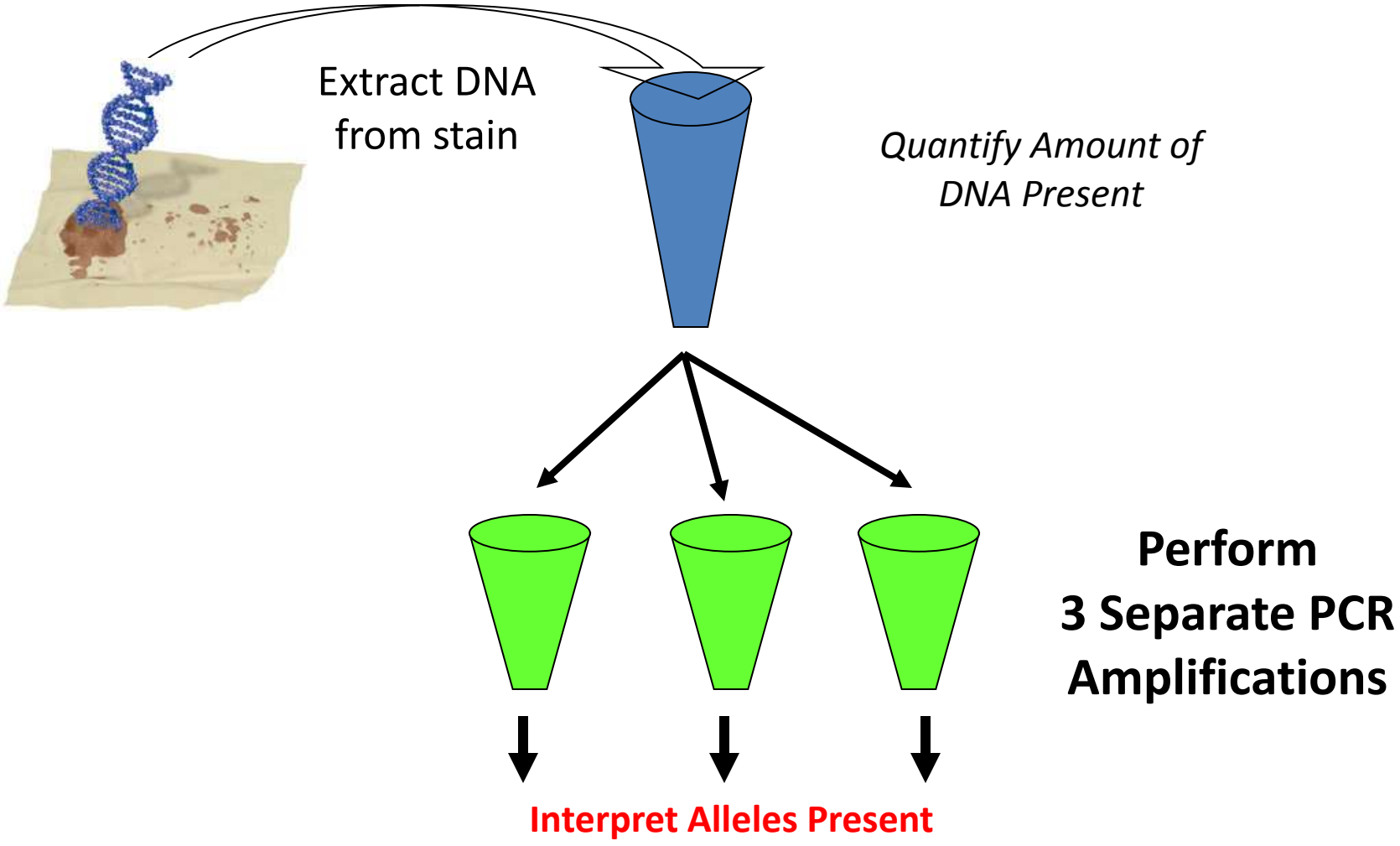
Identifiler, 10 pg DNA, 31 cycles

Allelic Drop-in



Identifiler, 10 pg DNA, 31 cycles

Typical LT-DNA Analysis Procedure

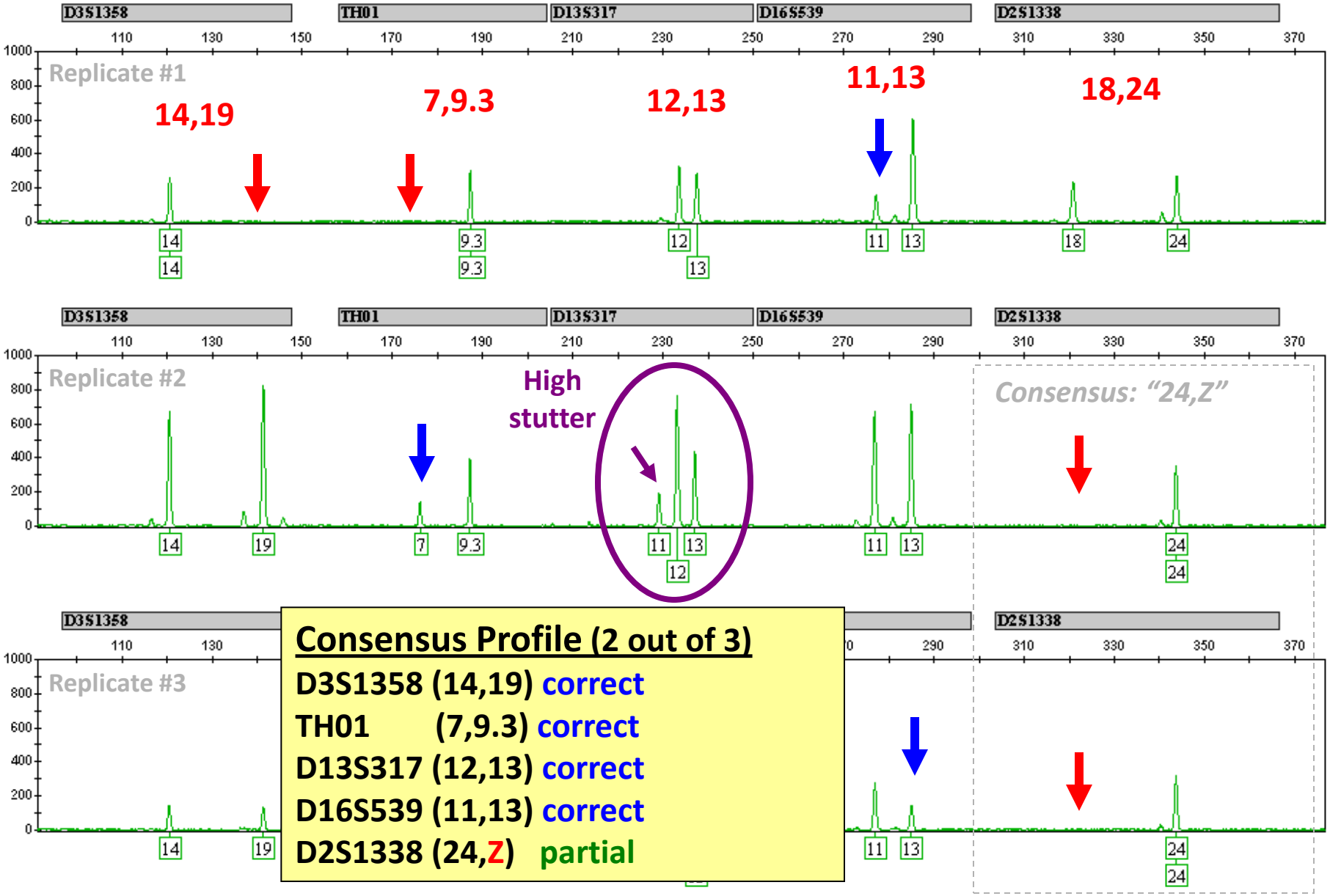


Suggestions for Optimal Results with LT-DNA

- Typically at least 2 – 3 PCR amplifications from the same DNA extract are performed to obtain consensus profiles
- An allele cannot be scored (considered real) unless it is present at least twice in replicate samples
- Extremely sterile environment is required for PCR setup to avoid contamination from laboratory personnel or other sources

10 pg template DNA with 31 cycles of PCR - triplicates

Identifiler data (green loci)



Consensus Profile (2 out of 3)

D3S1358 (14,19) correct
TH01 (7,9.3) correct
D13S317 (12,13) correct
D16S539 (11,13) correct
D2S1338 (24,Z) partial

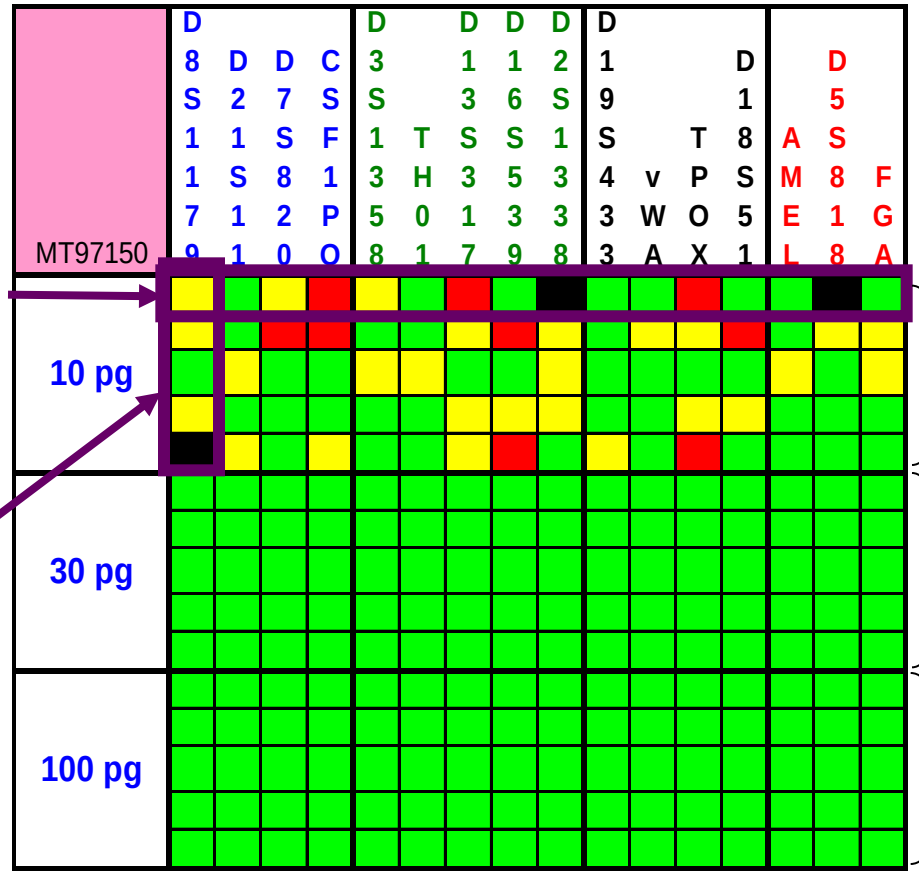
Allele PHR imbalance

Allele dropout

Sensitivity Comparison

Green = full (correct) type
Yellow = allele dropout
Red = locus dropout
Black = drop-in

Tested sample is heterozygous
(possesses 2 alleles) at every locus, which permits an examination of allele dropout



A single profile slice

A replicate slice

Results broken down by locus

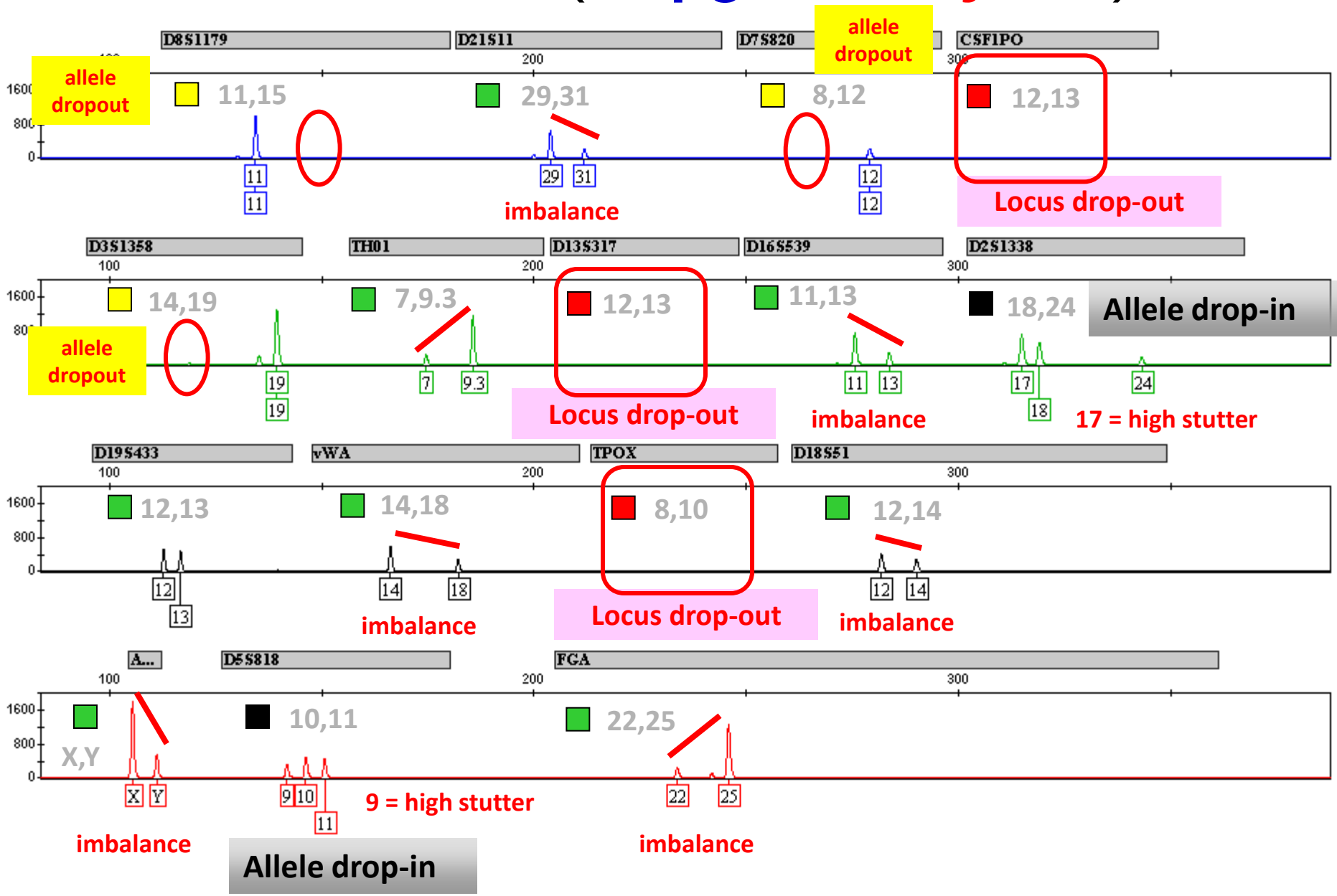
5 replicates of each [DNA]

5 replicates of each [DNA]

5 replicates of each [DNA]

10 pg (~2 cells)
30 pg (~6 cells)
100 pg (~18 cells)

Identifiler Plus (10 pg @ 32 cycles)

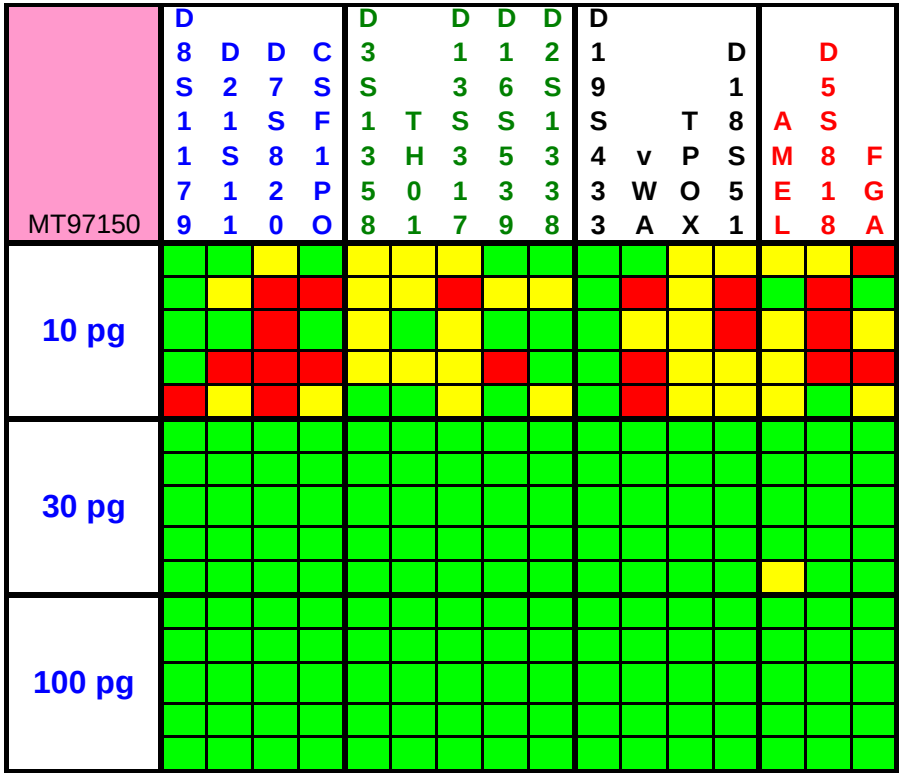


Impact of Three More PCR Cycles

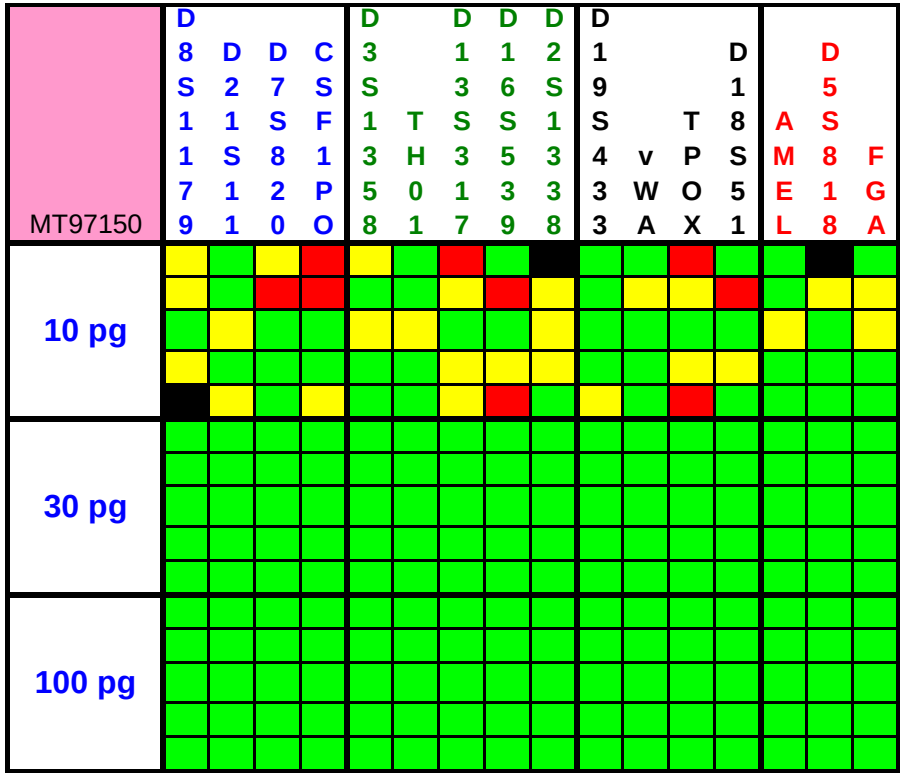
Identifiler Plus

Green = full (correct) type
Yellow = allele dropout
Red = locus dropout
Black = drop-in

*Full type = both alleles above 50 RFU
(does not account for peak imbalance)



29 Cycles



32 Cycles

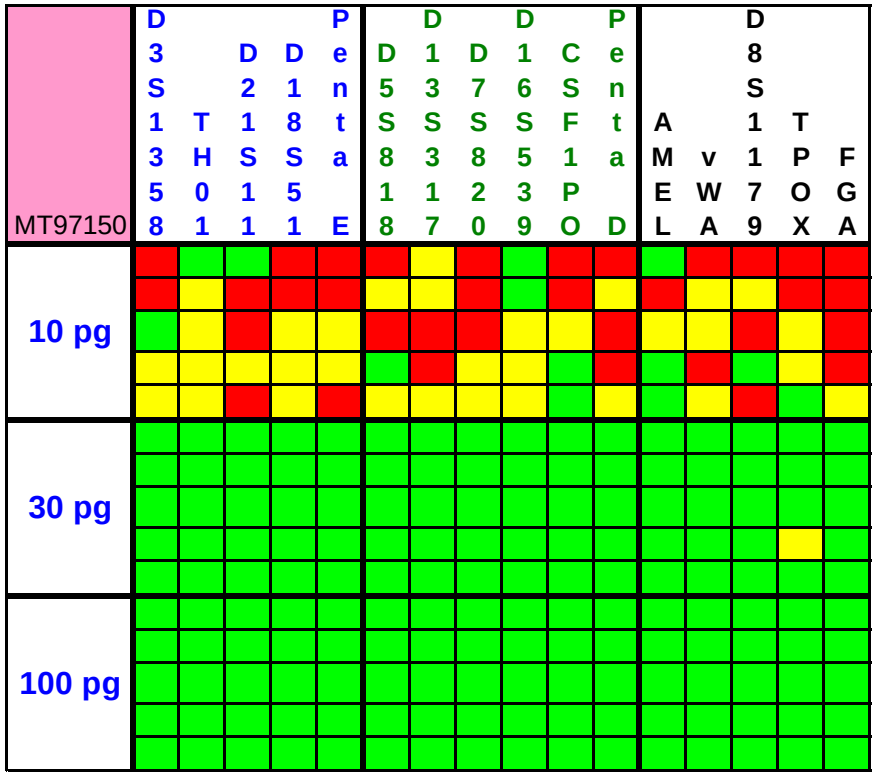
33% vs. 53% full profiles

38% improvement with 3 extra cycles

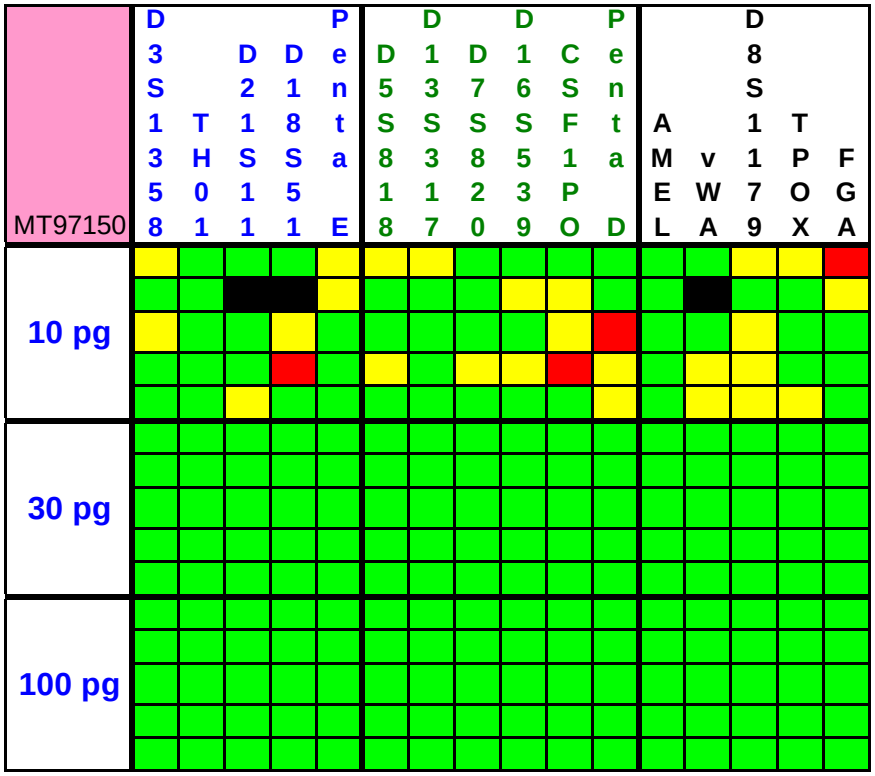
Sensitivity & Performance

PowerPlex 16 HS

Green = full (correct) type
Yellow = allele dropout
Red = locus dropout
Black = drop-in



31 Cycles



34 Cycles

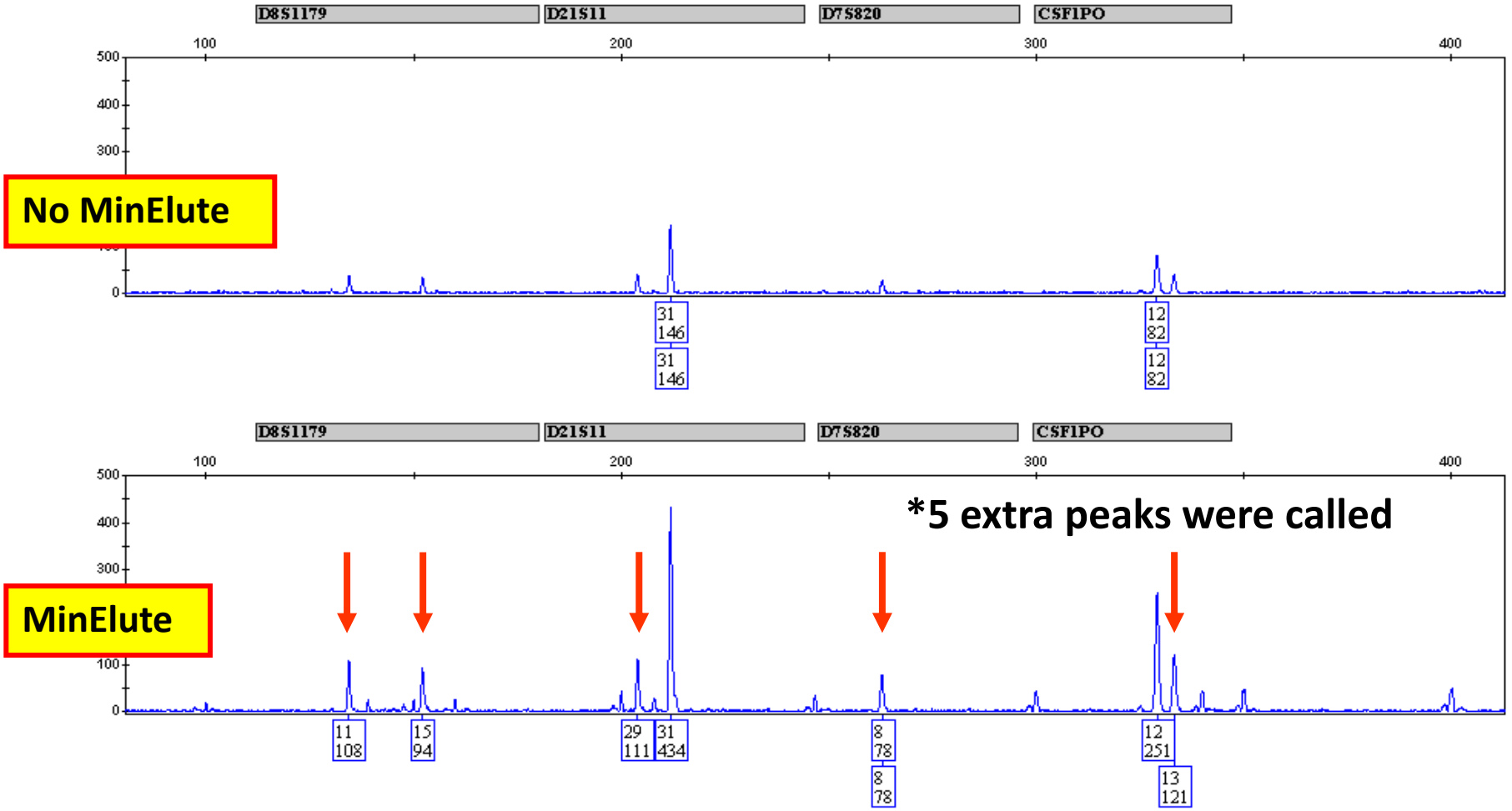
16% vs. 60% full profiles

73% improvement with 3 extra cycles

MinElute PCR Purification Kit

*96 well plates with vacuum protocol used

Identifiler Plus, 29 cycles, 10 pg



Signal Improvement:	~66%	~67%

Final Thoughts and Advice

Support to the Community

…Bringing traceability and technology to the scales of justice…

- Conduct interlaboratory studies
- Perform beta-testing of new human identity testing products
- **We collaborate with other NIJ grantees**
- We provide input to (or have aided):
  - Scientific Working Group on DNA Analysis Methods (**SWGDAM**)
  - Department of Defense Quality Assurance Oversight Committee for DNA Analysis
  - Virginia DFS Science Advisory Committee
  - American Prosecutor's Research Institute (**APRI**) DNA Forensics Program "Course-in-a-Box" for training lawyers
  - WTC Kinship and Data Analysis Panel (**KADAP**) and Hurricane Katrina efforts
  - NIJ Expert System Testbed (**NEST**) Project

A Few Words of Advice

- Hard work and studying are of the utmost importance in any field and really determining where your interests lie
- George Mason is an excellent school that opens many professional doors, especially in this area
- Making contacts and having professional relationships in the field is crucial to getting your foot in the door, but the rest is up to you!
- Having skills in speaking and writing is essential in this field
  - DNA analysts: writing reports and going to court
  - Research scientists: writing journal articles and giving presentations about your research

A Few Useful Websites:

- U.S. Federal Government jobs
  - www.usajobs.gov

- American Academy of Forensic Sciences
  - www.aafs.org

- Mid-Atlantic Association of Forensic Scientists
  - www.maafs.org

- National Institute of Justice
  - www.dna.gov

- **STRBase**
  - www.cstl.nist.gov/strbase

Thank you for your attention

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

Contact Information

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

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

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