FORENSICS @ NIST
November 28-30, 2012 • #NISTForensics

STR Loci and Multiplex Kits

Becky Hill

Research Biologist, Applied Genetics Group

Forensics@NIST 2012 Meeting
Gaithersburg, MD
November 28, 2012

Outline of Topics to Discuss

- Brief background on STR loci and kits
  - What are they and why are they important?
- The role of NIST in STR typing
  - NIST unrelated 1036 U.S. population samples
  - Concordance testing
- New STR multiplex kits available
  - GlobalFiler (Life Technologies)
  - PowerPlex Fusion (Promega)
  - PowerPlex Y23 (Promega)

Background on STR Loci and Kits

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

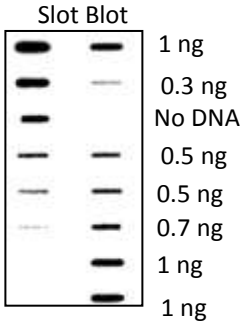
Blood Stain

Buccal swab

Sample Collection & Storage

Biology

DNA Extraction

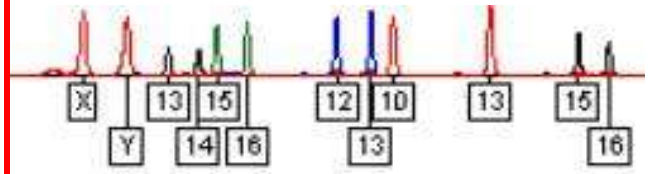


DNA Quantitation

Multiplex PCR Amplification

Technology

DNA separation and sizing



STR Typing

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

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

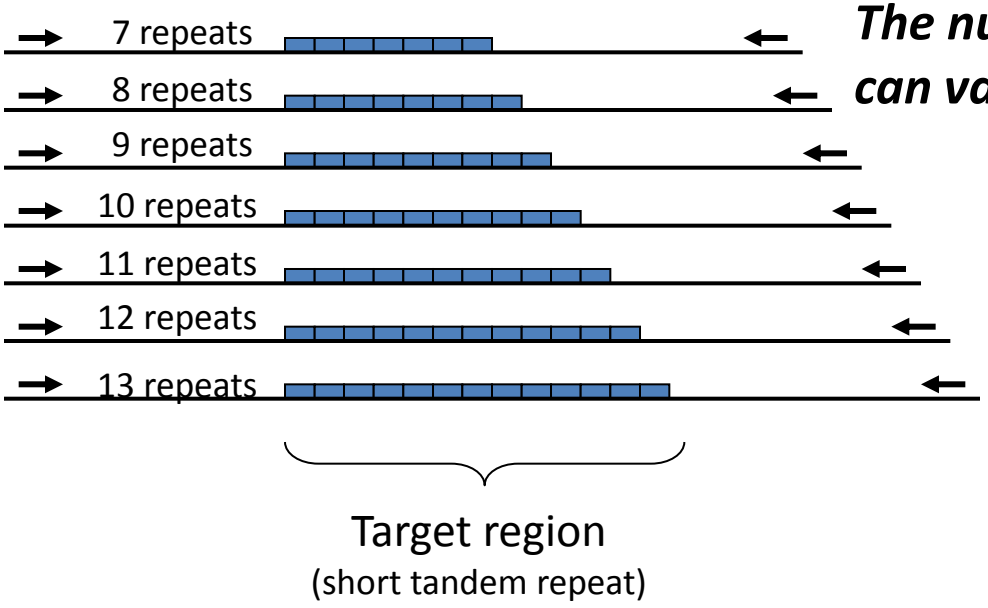
Interpretation of Results

Short Tandem Repeat (STR) Markers

An accordion-like DNA sequence that occurs between genes

TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGAAGACAGGTG**GATAGATA**
GATAGATAGATAGATAGATAGATAGATAGATAGATAGATATCATTGAAAGACAAA
ACAGAGATGGATGATAGATACATGCTTACAGATGCACAC

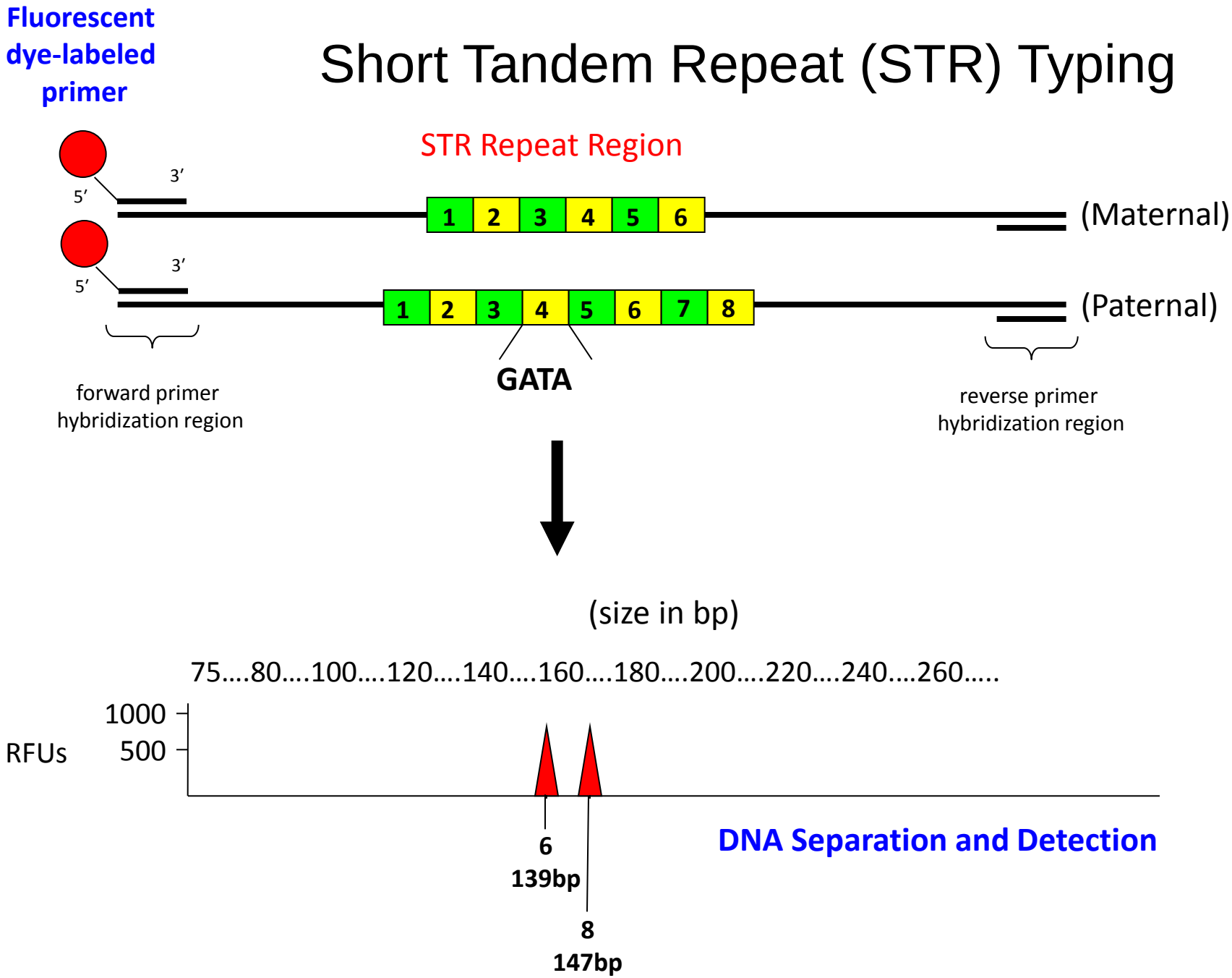
= 12 GATA repeats ("12" is all that is reported)



The number of consecutive repeat units can vary between people

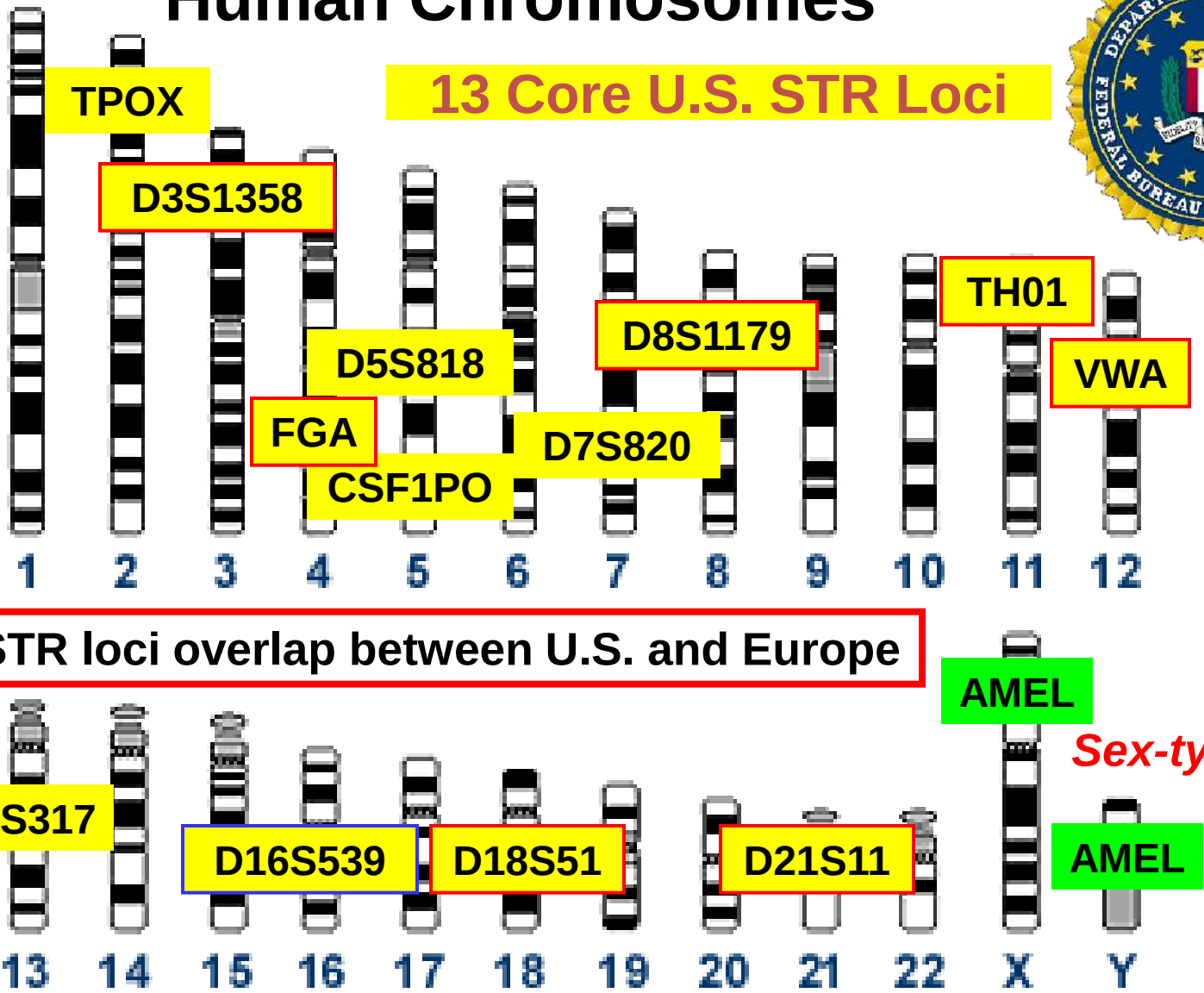
The FBI has selected **13 core STR loci** that must be run in all DNA tests in order to provide a common currency with DNA profiles

Short Tandem Repeat (STR) Typing



NIST Applied Genetics

Position of Forensic STR Markers on Human Chromosomes



Commercially Available STR Kits

Applied Biosystems (18)

- ~~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)
- **GlobalFiler** (2012)

Promega Corporation (17)

- 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 CS7 (2009)
- PowerPlex 18D (2011)
- PowerPlex Y23 (2012)
- PowerPlex 21 (2012)
- **PowerPlex Fusion** (2012)

Qiagen (10) kits in 2010

Primarily selling kits in Europe

Due to patent restrictions cannot sell in U.S.

Investigator kits

- ESSplex
- ESSplex SE
- Decaplex SE
- IDplex
- Nonaplex ESS
- Hexaplex ESS
- HDplex
- Triplex AFS QS
- Triplex DSF
- Argus X-12

>1/3 of all STR kits were released in the last four years

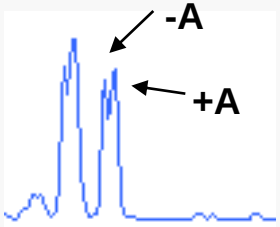
Impact of DNA Amount into Multiplex PCR Reaction

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)



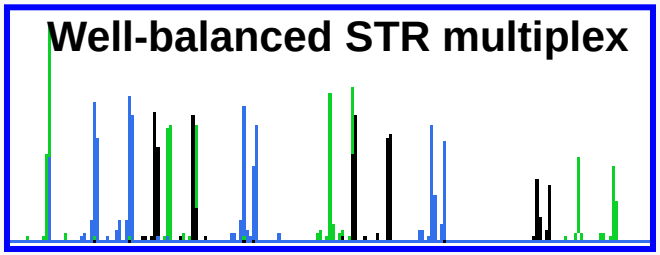
Too much DNA

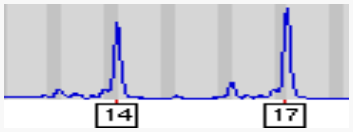
- Off-scale peaks
- Split peaks (+/-A)
- Locus-to-locus imbalance

2.0 ng

STR Kits Work Best in This Range

0.5 ng

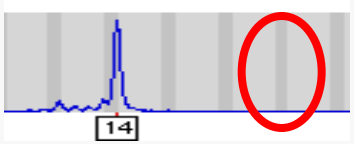




100 pg template

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

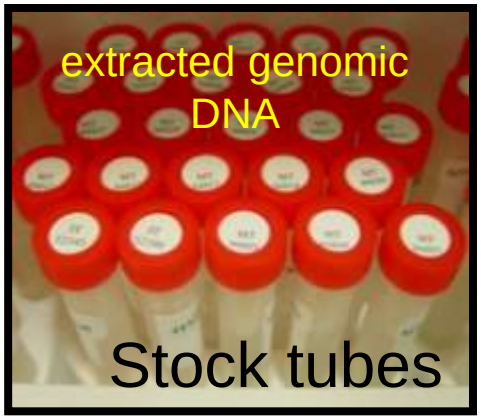
NIST U.S. Population Samples

NIST Standard Sample Sets

- **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)

>1450 total samples

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



NIST 1036 U.S. Population Samples

- 1032 males + 4 females
  - 361 Caucasians (2 female)
  - 342 African Americans (1 female)
  - 236 Hispanics
  - 97 Asians (1 female)

***Unrelated* samples**
All known or potential related individuals (based on autosomal & lineage marker testing) have been removed from the 1036 data set (e.g., only sons were used from father-son samples)

- Anonymous donors with self-identified ancestry
  - Interstate Blood Bank (Memphis, TN) – obtained in 2002
  - Millennium Biotech, Inc. (Ft. Lauderdale, FL) – obtained in 2001
  - DNA Diagnostics Center (Fairfield, OH) – obtained in 2007

- **Complete profiles with 29 autosomal STRs + PowerPlex Y23**
  - **Examined with multiple kits** and in-house primer sets enabling concordance

- Additional DNA results available on subsets of these samples
  - mtDNA control region/whole genome (AFDIL)
  - >100 SNPs (AIMs), 68 InDel markers, X-STRs (AFDIL)
  - NIST assays: miniSTRs, 26plex, >100 Y-STRs, 50 Y-SNPs

Data available on STRBase: **http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm**

NIST
Applied Genetics

Benefits of NIST 1036 Data Set

- **Elimination of potential null alleles due to primer binding site mutations** through extensive concordance testing performed with different PCR primer sets from all available commercial STR kits
- **Ancestry testing performed** on DNA samples with autosomal SNPs, Y-SNPs, and mtDNA sequencing to verify self-declared ancestry categorization
- **Related individuals removed** based on Y-STR and mtDNA results
- **Full characterization** of all commercial STR loci based on population statistics

Characterizing New STR Loci

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 29 commonly used autosomal STR loci and 23 Y-STR loci

Presentations/Publications:

- Hill et al (2011) *FSI Genetics* 5(4): 269-275
- Hares (2012) Expanding the U.S. core loci… *FSI Genetics* 6(1): e52-e54
- Butler & Hill (2012) *Forensic Sci Rev* 24(1): 15-26
- Hill et al (2012) poster at ISHI: http://www.cstl.nist.gov/biotech/strbase/pub_pres/Hill-ISHI2012-STRloci.pdf

NIST U.S. Population Data

- The data from our 1036 U.S. population samples is now available on STRBase:

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

- An summary of the NIST 1036 data set was recently published in Profiles in DNA for autosomal and YSTR loci
- Population data have been submitted to *FSI: Genetics* for publication
  - 29 autosomal STR loci (*Hill et al*)
  - 23 Y-STR loci (*Coble et al*)

Concordance Evaluation of STR Kits

Working with Forensic Commercial Companies

- Work primarily with 3 forensic commercial companies
  - Applied Biosystems – Foster City, CA
  - Promega – Madison, WI
  - Qiagen - Germantown, MD (parent company in Hilden, Germany)
- What is provided to NIST
  - Prototype DNA kits for evaluation with NIST standard U.S. samples
- What we provide to the companies
  - Complete data evaluation and publications of findings with their approval

Importance of Data Comparisons Between DNA Kits

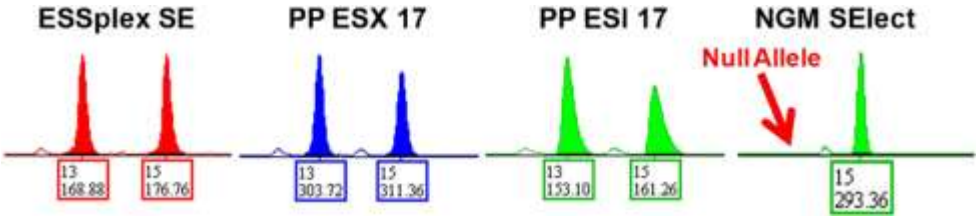
- There are a variety of commercial DNA kits with different configurations of markers
  - Differences in profiles are rare, but can and do occur

- **Discordant results can impact DNA databases**
  - Currently ~11.5 million profiles in the national database
  - Information sharing occurs between state and local databases
  - If there are differences between kits, this can be detrimental (could lead to false negatives)

- Concordance with NIST reference materials is valuable for proper calibrations of all kits used
  - All forensic labs are required to test NIST SRM 2391c per FBI QAS
  - Important to test with all new DNA kits to determine and characterize any differences

STR Kit Concordance Studies

D18S51 Comparisons



D18S51 null allele with the NGM SElect kit as compared to the ESSplex SE kit, PowerPlex ESX 17 and ESI 17 systems

*Kits are kindly provided by **Applied Biosystems, Promega, and Qiagen** for concordance testing performed at NIST*


Forensic Science International: Genetics Supplement Series 3 (2011) e188–e189

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

ELSEVIER

Concordance testing comparing STR multiplex kits with a standard data set

Carolyn R. Hill *, Margaret C. Kline, David L. Duewer, John M. Butler

U.S. National Institute of Standards and Technology, NIST 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA

- Examined NIST samples across >20 STR kits and in-house assays covering 29 autosomal STR loci
- **99.90% concordance observed to-date**
  - 1,225 total differences due to primer binding site mutations from 1,176,994 allele comparisons (as of Nov 2012)
- Information provided back to kit developers to redesign primers or add extra ones – often prior to kit release

Benefits to the Forensic Community

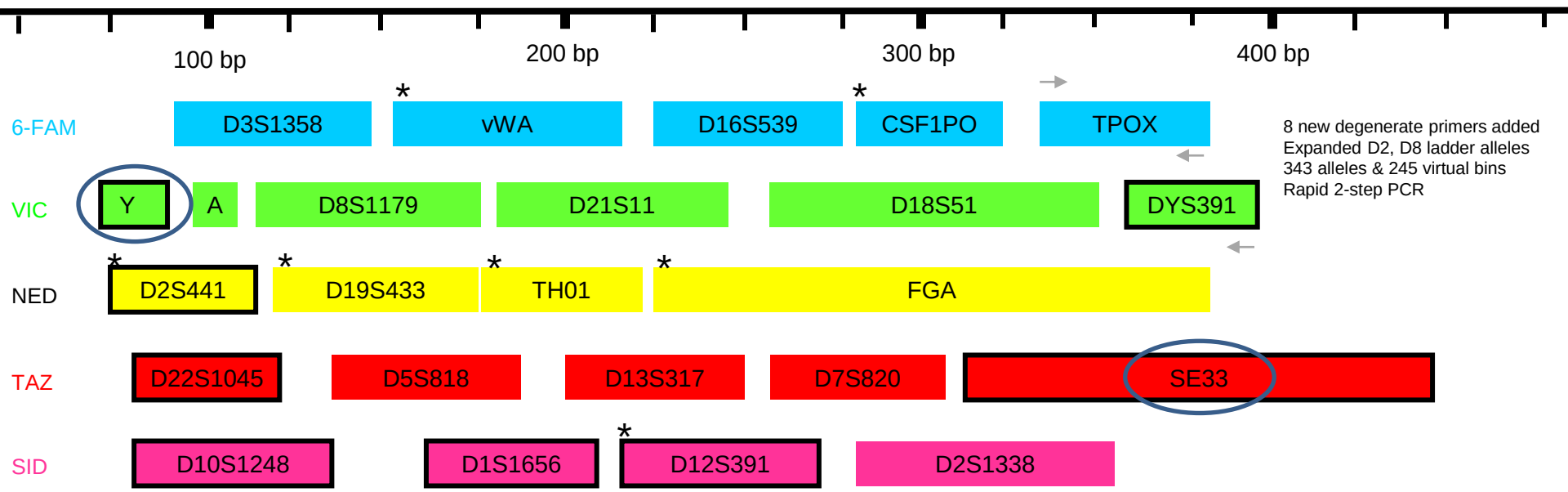
- **With NIST input and testing, companies can fix their kits before they are released and allow them get to the market faster with fewer issues**
- The market share for these DNA analysis systems (instrumentation and reagents) is in the $150-200 M per year
  - NIST has a direct impact by helping to improve the technology
- NIST has a neutral and confidential approach to working with different commercial companies simultaneously
- NIST has access to a useful set of standard U.S. population samples for these kit evaluations
- NIST publications in peer-reviewed journals and presentations in domestic and international conferences and webinars

New STR Multiplex Kits
Recently Launched

STR Kit Layouts by Dye Label and PCR Product Size

Life Technologies/Applied Biosystems **GlobalFiler** (6-dye – LIZ600 size standard)

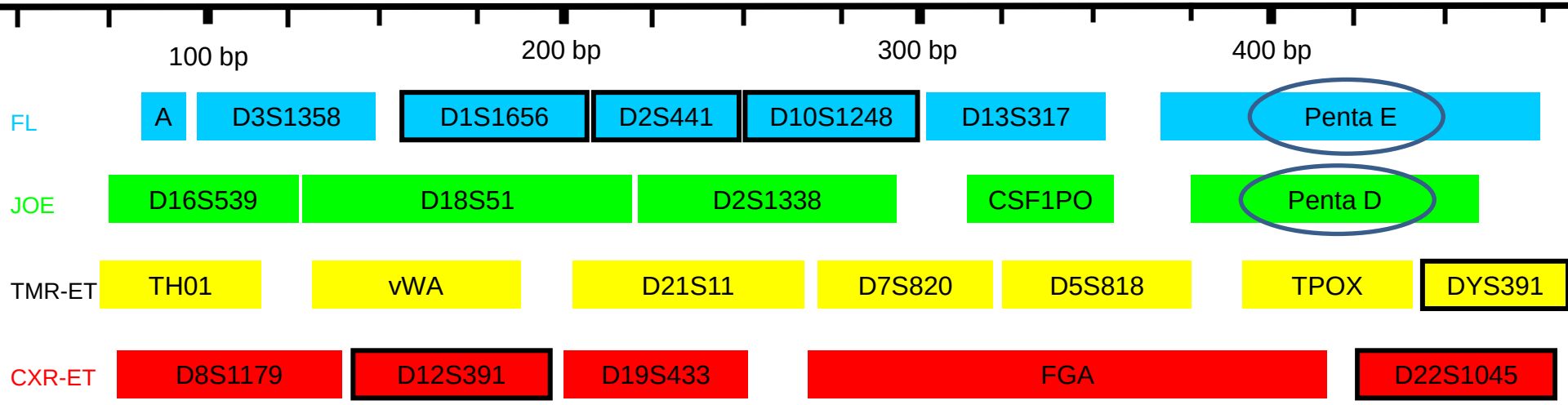
24plex



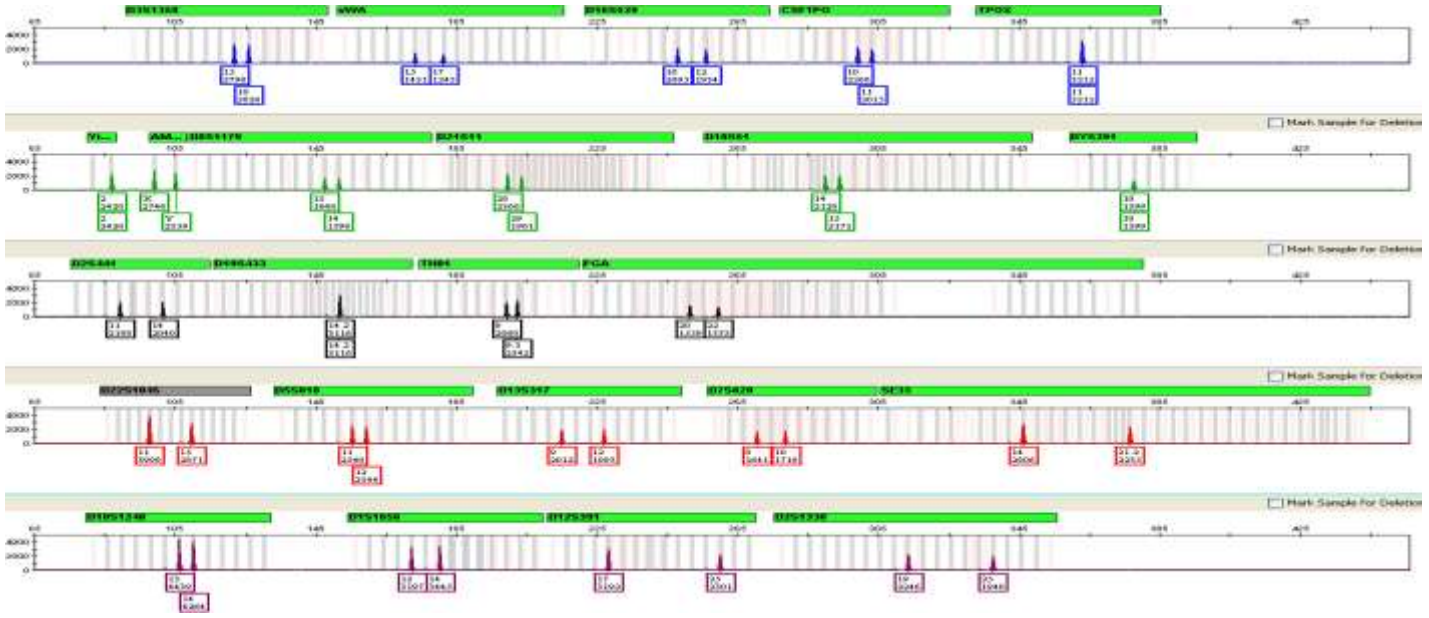
8 new degenerate primers added
Expanded D2, D8 ladder alleles
343 alleles & 245 virtual bins
Rapid 2-step PCR

Promega PowerPlex **FUSION** (5-dye – CC5 internal lane standard 500)

24plex

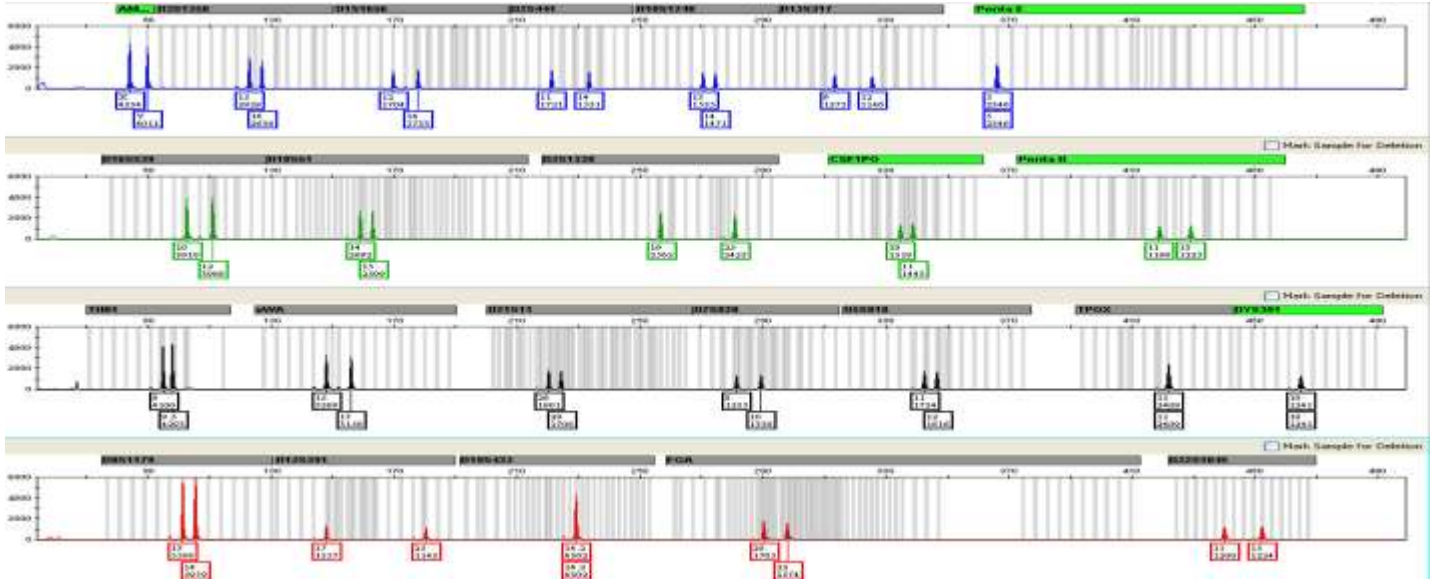


Applied Biosystems **GlobalFiler**

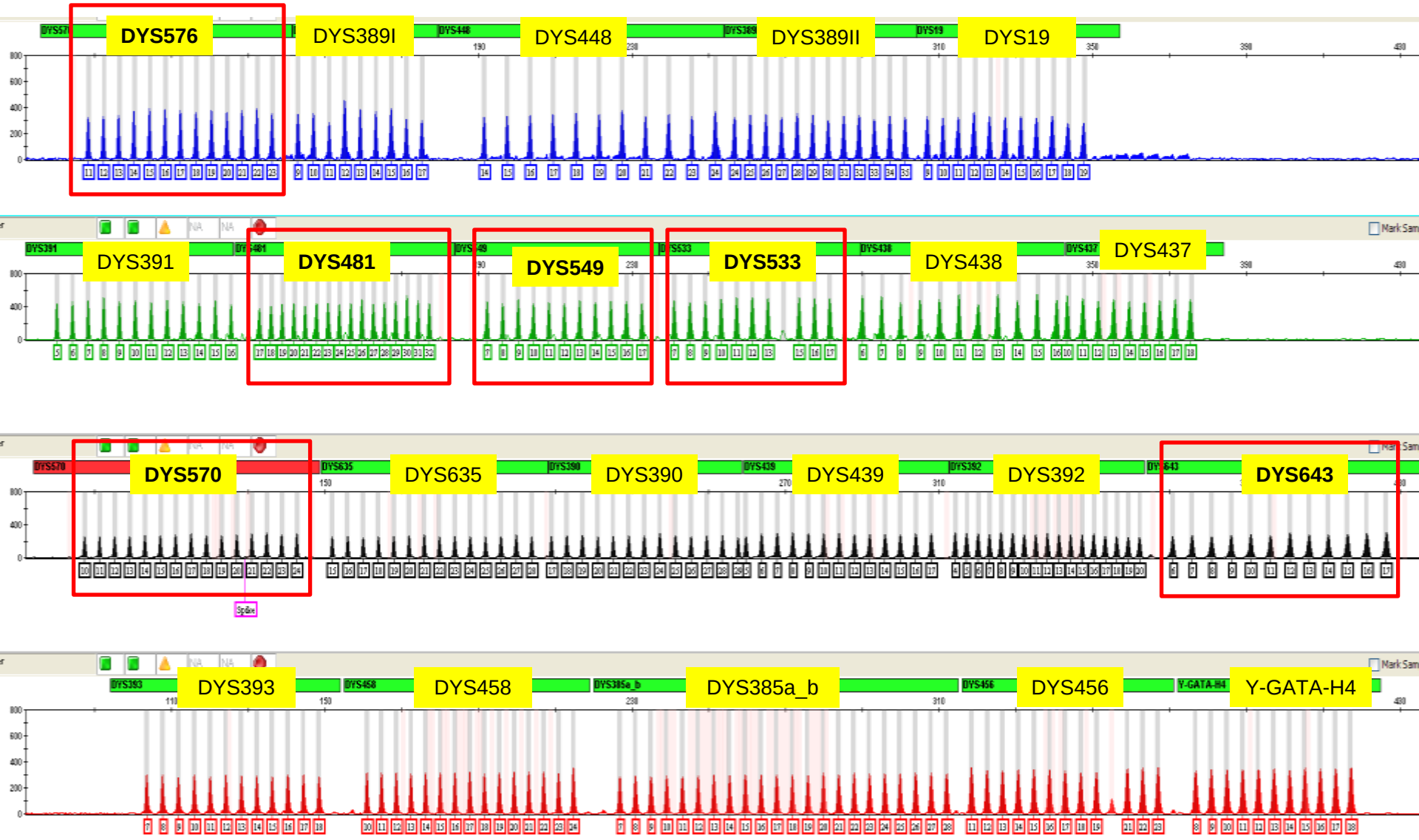


Both kits were released in Sept 2012

Promega PowerPlex **Fusion**



PowerPlex Y23 Allelic Ladders



Summary

- NIST has a set of 1036 U.S. population samples that have been used to fully characterize 29 autosomal STR loci and 23 Y-STR loci
- NIST plays an important role in concordance testing to aid the community
  - Several null alleles have been fixed before the final release of new STR multiplex kits
- Commercial companies are continuing to release larger STR multiplexes to meet the needs of the forensic community

NIST Applied Genetics

Acknowledgments

NIST Team for This Work


John Butler


Dave Duewer


Margaret Kline


Mike Coble

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

A special thanks to Applied Biosystems, Promega, and Qiagen for providing the kits used in this study

Contact Info:
becky.hill@nist.gov
301-975-4275

NIJ National Institute of Justice