NIST Standards for **Genetic Testing**: Past, Present, and Future

John M. Butler, Ph.D.
National Institute of Standards and Technology
Biochemical Science Division/Applied Genetics Group

SACGHS
December 1, 2008

Presentation Overview

- Past
  - Extensive experience with developing forensic DNA reference materials and genotyping assays and technologies
- Present
  - Applied Genetics Group to consolidate forensic DNA with clinical genetics and agricultural biotech efforts
  - Work with genetic genealogy
- Future
  - Planned genetic testing standards

Congress Passed **the DNA Identification Act of 1994** (Public Law 103 322)

Formalized the FBI's authority to establish a national DNA index for law enforcement purposes.

FBI's DNA Advisory Board

Quality Assurance Standards for Forensic DNA Testing Laboratories

(October 1, 1998)

STANDARD 9.5

The laboratory shall check its DNA procedures annually or whenever substantial changes are made to the protocol(s) against an appropriate and available NIST standard reference material or standard traceable to a NIST standard.

Checks and Controls on Forensic DNA Results

Community	FBI DNA Advisory Board's Quality Assurance Standards *(also interlaboratory studies)*
Laboratory	ASCLD/LAB Audits and Accreditation
Analyst	Proficiency Tests & Continuing Education
Method/Instrument	Validation of Analytical Performance *(with aid of **traceable reference materials**)*
Protocol	Standard Operating Procedure is followed
Data Sets	Allelic ladders, positive and negative amplification controls, and reagent blanks are used
Individual Sample	Internal size standard present in every sample
Interpretation of Result	Second review by qualified analyst/supervisor
Court Presentation of Evidence	Defense attorneys and experts with power of discovery requests

The Tools of DNA Typing and SRM Needs

- RFLP Testing (Late 1980's) ~~SRM 2390~~
  - Radioactive Based
  - Chemiluminescent Based

  Technology no longer used
- PCR-Based Testing (Mid 1990's)
  - Dot-Blot
  - VNTR
  - STR (Fluorescent markers used today)

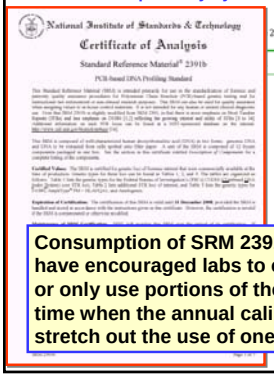
  SRM 2391..a..b

  Growth area
- DNA Sequencing (Late 1990's) SRM 2392, 2392-I
  - Mitochondrial DNA
- Y-Chromosome Testing (early 2000's) SRM 2395

  Growth area

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

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					
	9,10	11,12					
	6,9	10,13					
	8,9	11,13					
	9,10	11,11	10,12	9,12	12,14	23,25	12,14
	6,8	11,11*	11,12	3.2,13	12,16	17,22	13,15.2

22 autosomal STRs characterized across 12 DNA samples

Consumption of SRM 2391b has slowed because we have encouraged labs to create NIST-traceable materials or only use portions of the SRM's 12 components each time when the annual calibrations are performed (i.e., to stretch out the use of one unit of SRM 2391b)

Steps in Forensic 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

DNA Extraction

DNA Quantitation

Biology

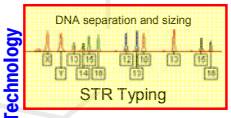
Multiplex PCR Amplification

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

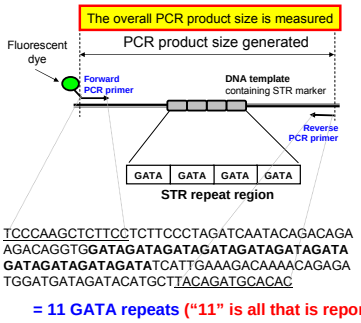


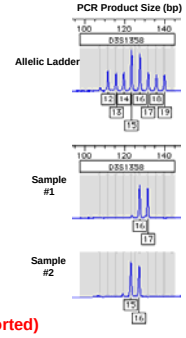
Male: 13,14-15,16-12,13-10,13-15,16

Interpretation of Results

Short Tandem Repeat (STR) Markers

PCR primers anneal to unique sequences bracketing the variable STR repeat region





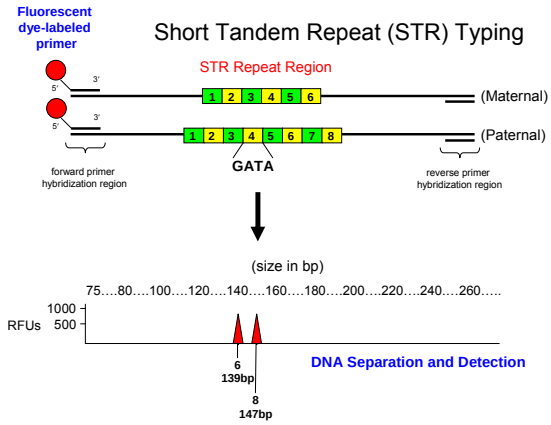
= 11 GATA repeats ("11" is all that is reported)

Position of Forensic STR Markers on Human Chromosomes

13 Core U.S. STR Loci

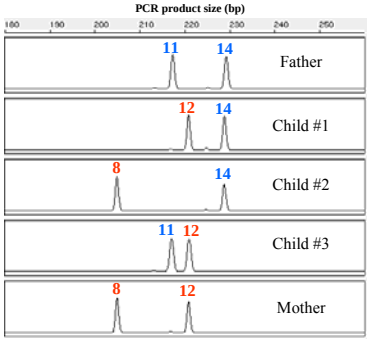
8 STR loci overlap between U.S. and Europe

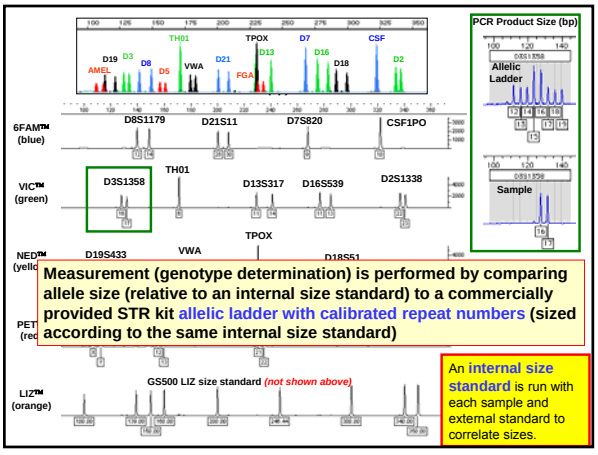
Short Tandem Repeat (STR) Typing



PATERNITY TESTING

Family Inheritance of STR Alleles (D13S317)





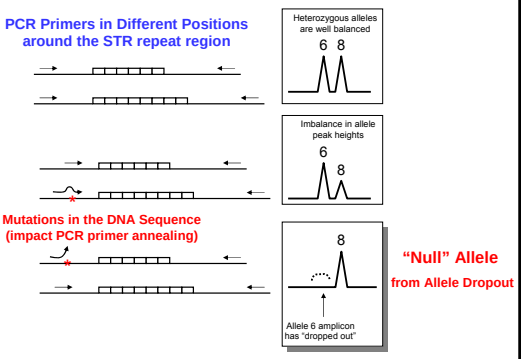
Measurement (genotype determination) is performed by comparing allele size (relative to an internal size standard) to a commercially provided STR kit allelic ladder with calibrated repeat numbers (sized according to the same internal size standard)

An **internal size standard** is run with each sample and external standard to correlate sizes.

Different Genetic Tests Can Give Different Results Based on PCR Primer Positions

PCR Primers in Different Positions around the STR repeat region

Mutations in the DNA Sequence (impact PCR primer annealing)



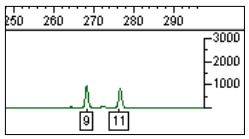
"Null" Allele from Allele Dropout

Figure 6.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

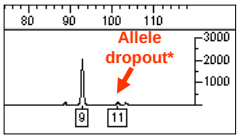
SRM 2391b Genomic 8 with D16S539

All allele calls with MiniFiler for CSF1PO, D7S820, D13S317, D18S51, D21S11, FGA, and D16S539 (with the exception noted below) ***match previously certified values.***

Identifiler

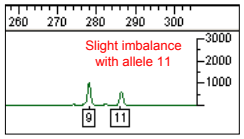


MiniFiler



*Due to primer binding site mutation

PowerPlex 16



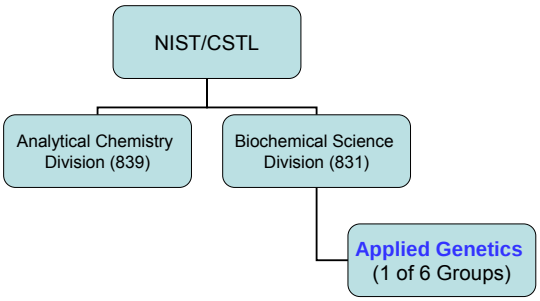
National Institute of Justice

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

Current Areas of NIST Effort with Forensic DNA

- **Standards** http://www.cstl.nist.gov/biotech/strbase/
  - **Standard Reference Materials**
  - Standard Information Resources (STRBase website)
  - Interlaboratory Studies
- **Technology**
  - Research programs in SNPs, miniSTRs, Y-STRs, mtDNA, qPCR
  - Assay and software development
- **Training Materials**
  - Review articles and workshops on STRs, CE, validation
  - PowerPoint and pdf files available for download

An Abbreviated Organizational Chart



Responsibilities for Forensic DNA Testing and Clinical and Agricultural Diagnostics (GMOs)

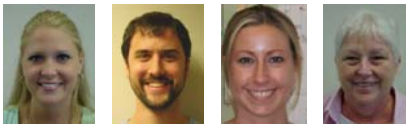
NIST Applied Genetics Group

Applied Genetics

Formally organized October 2008

Group Leader

John Butler | Marcia Holden | Margaret Kline | Pete Vallone



Amy Decker | Ross Haynes | Becky Hill | Jan Redman

Group Mission Statement

Applied Genetics

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

- **forensic DNA testing,**
- **clinical genetics,**
- **agricultural biotechnology,** and
- **DNA biometrics.**

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)
- NIST (SRM development and production)

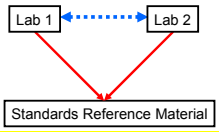
We are looking to strengthen our portfolio in clinical genetics and agricultural biotech

Standard Reference Materials (SRMs)

http://www.nist.gov/srm

Traceable standards to ensure accurate and comparable measurements between laboratories

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

NIST Applied Genetics

NIST DNA Reference Materials

Date of release or certificate revision (r)

Forensic Applications

- **STR PCR DNA Profiling** (SRM 2391b) – 1995, r2008
- Mitochondrial DNA Sequencing (SRM 2392-I, 2392) – 1999, 2003
- **Human Y-Chromosome DNA Profiling** (SRM 2395) – 2003, r2008
- RFLP DNA Profiling (SRM 2390) – 1992, r2001, *now obsolete*

Clinical Applications

- Fragile X Human DNA Triplet Repeat (SRM 2399) – 2004, r 2007
- Huntington's Disease CAG Repeats (SRM 2393) – in process

Platform Testing

- **Human DNA Quantitation** (SRM 2372) - 2007
- Heteroplasmic mtDNA Mutation Detection (SRM 2394) - 2004
- DNA Sequence Library for External RNA Controls (SRM 2374)

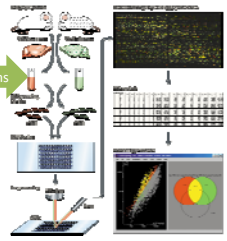
A few others are in early stages of development

Slide from **Marc Salit**, Multiplexed Biomolecular Science Group

External RNA Control Consortium

- Industry-initiated, NIST-hosted, stakeholder coupled
  - Janet Warrington, VP Clinical Genomics at Affymetrix
  - all major microarray technology developers
  - other gene expression assay developers
    - collaborative study
    - probe content on commercial array platforms
- Use reference material approach to transfer accuracy of NIST measurements and ensure harmony amongst users
  - Long-term useful for gene expression, not tied to microarray measurement approach
- Novel aspects
  - Certification of sequence
    - developing new metrological framework for certifying sequence as property, consistent with ISO/REMCO definition of CRM
    - focus on confidence in sequence
  - SRM to be template; work with SDO to develop documentary standard for CRM production
    - CRMs to be commercially available
- Model for future work in this area
  - sets up our work to be well-coupled to stakeholder needs
  - keeps us relevant and tied in
  - lets us develop SRMs that are stable and generic

Spike-ins

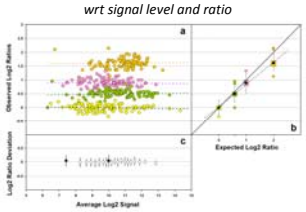


Slide from **Marc Salit**, Multiplexed Biomolecular Science Group

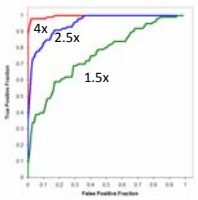
RNA Control Set

enable objective performance measures for microarray gene expression

Analytical performance
wrt signal level and ratio



Receiver-Operating Characteristic Curves



Approach developed in conjunction with Pine and Thompson, FDA
– results of simulation shown –

Slide from **Marc Salit**, Multiplexed Biomolecular Science Group

NIST SRM 2374 – DNA Sequence Library for External RNA Controls

- NIST developing reference material of 96 control sequences
  - SRM will be plasmid DNA with control sequences as inserts
  - sequence is certified property
  - sequencing at NIST and multiple partner labs
  - sequencing with Sanger and next-gen "UHTS" approach(es)
- Developed sequence library from submission by ERCC members & synthesis
  - evaluated performance of RNA controls on variety of platforms
  - selected 96 well performing sequences
- Preparing SRM
  - cloned sequence library into common vector
    - suitable for use in accurate preparation of RNA controls
  - Prepared 400 units
    - 96 tubes in each
- Certifying ~100,000 bases
  - Sanger sequencing complete at CBI, NIST
  - alternate sequencing approaches underway
  - quality measures developed to permit estimation of sequence reliability
    - based on *de novo* assembly at alternate sites
    - integration of data from multiple labs

Two Different Nomenclatures Used by DNA Ancestry Testing Companies for Y-chromosome Marker DYS442

(A) TATTCCATTG **TATC TATC *TGTC TGTC TGTC***
TATC TATC TATC TATC TATC TATC TATC
TATC TATC TATC TATC TATC ACAGTTTCTT

$[TATC]_2$ $[TGTC]_3$ $[TATC]_{12}$ — Gusmão et al. (2006); ISFG recommended

NIST supports the (A) nomenclature

(B) TATTCCATTGTATCTATCTGTCTGTCTGTC **TATC**
TATC TATC TATC TATC TATC TATC TATC
TATC TATC TATC TATC ACAGTTTCTT

$[TATC]_{12}$ — Iida et al. (2001)

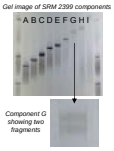
Butler et al. (2008) Addressing Y-STR allele nomenclature. *Journal of Genetic Genealogy* 4(2): 125-148

Fragile X (SRM 2399)

(now available)

NIST Applied Genetics

(revised certificate in 2007)



Assigned repeat values

A = 20	Normal
B = 30	
C = 41	
D = 51	Gray Zone
E = 60	
F = 73	
G = 88/89, 93	Premutation
H = 96	
I = 118	

Currently no full mutation components available

- **Disease:** **Fragile X involves the expansion of a trinucleotide (CGG) repeat** and is the most common inherited cause of intellectual impairment and the most common known genetic cause of autism that affects 1 in 4000 males and 1 in 6000 females of all races and ethnic groups. Normal (29-31 CGG repeats), Premutation (55-200 CGG repeats), Full Mutation (more than 200 CGG repeats), and Intermediate or Gray Zone Alleles (40 - 60 repeats) [1].
- **SRM 2399 Components**: Nine PCR products (labeled "A" to "I") and possessing between 20 and 118 CGG repeats (along with 220 bp of flanking region sequence) were initially released in Dec. 2004; revisions were made to the certificate in June 2007; component G was found to have two alleles present that were assigned to contain 88/89 and 93 CGG repeats.
- **Customer Feedback Requested**: We will consider adding components with larger number of repeats if there is interest.

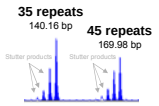
[1] Sherman, S. (2002). "Epidemiology". Chapter 3 in Fragile X Syndrome, Diagnosis Treatment and Research. Ed. Hagerman, R. J. & Hagerman, P.J. (3rd Edition). Johns Hopkins University Press: Baltimore.

Huntington's Disease (SRM 2393)

(in process)

NIST Applied Genetics

(completion planned for 2009)



Repeat count	Classification	Disease status
<27	Normal	Unaffected
27–35	Intermediate	Unaffected
36–39	Variable Penetrance	+/- Affected
>39	Full Penetrance	Affected
>60	Juvenile onset	Affected

- **Disease:** Huntington disease is a neurodegenerative disease of midlife onset that produces choreic movements and cognitive decline, often accompanied by psychiatric changes that affects approximately 1 in 10,000 individuals. Inheritance is autosomal dominant with clinical manifestations associated with **expansion of a polymorphic trinucleotide (CAG) repeat.**
- **Plan for Standards**: Supply DNA fragments that are larger than the currently published genotyping primers with components to have repeat numbers representing "normal" to "Juvenile onset" stages.
- **Potential Users:** Clinical diagnostic laboratories wanting to ensure the accuracy and comparability of their testing results to other testing laboratories and those wishing to validate their CAG repeat sizing methods.

AMERICAN COLLEGE OF MEDICAL GENETICSStandards and Guidelines for Clinical Genetics Laboratories 2006 Edition Technical Standards and Guidelines for Huntington Disease:

Clinical Guidelines for Huntington's Disease

Normal	Mutable Normal *Intermediate alleles*	HD Alleles Reduced penetrance	HD Alleles Full penetrance
≤ 26 CAG repeats	27 – 35 CAG repeats	36 – 39 CAG repeats	≥ 40 CAG repeats

Recommended HD sizing accuracy:
± 1 repeat for alleles ≤ 43
± 2 repeat for alleles between 44 -50
± 3 repeat for alleles between 51 -75
± 4 repeat for alleles > 75

American College of Medical Genetics: **Standards and Guidelines for Clinical Genetics Laboratories** *2006 Edition* : **Technical Standards and Guidelines for Huntington's Disease**

Cytomegalovirus (CMV) Reference Material

(in development)

NIST Applied Genetics

(completion planned for 2010)

- **Disease:** CMV causes life-threatening infections in immunocompromised patients and in congenital transmission to infants, though commonly found and usually latent in the general population
- **Standard needs:** Calibration and quality control of quantitative real-time PCR assays of blood and other body fluids
- **Customers:** Producers of secondary standards or clinical laboratory using in-house assays for CMV detection
- **Current Plans:**
  - ***Materials***: Pure viral DNA from Towne Strain in the form of a bacterial artificial chromosome (Towne ΔUL147) containing all of the viral genome except for region US1-15 and UL147; viral DNA to be provided in a buffer for dilution into a user's matrix of choice
  - ***Certification***: the mass fraction of DNA based on SI traceable phosphorus measurements
  - ***Additional Information***:
    - Copy number from DNA mass fraction and digital PCR
    - DNA sequence of genes that are targets for Q-PCR
    - Restriction digest and microarray validation of BAC
    - Testing of various published PCR assays
    - Homogeneity and stability

Agricultural Biotechnology

(in development)

NIST Applied Genetics

Biotech Crop Quantification (grain/food)

- **Background**: Shipments of US grain/food are tested for the type and quantity of biotech crop material to comply with regulatory policies of importing countries. Multiple testing methods often give different results. NIST is working with biotech crop developer companies to provide a quantitative real-time PCR method that would be applicable to numerous biotech crop events.
- **Strategy:** Detection and quantification of the 35S promoter sequence from the Cauliflower mosaic virus.
  - The 35S promoter is a strong and consistent promoter sequence and has been used in genetic constructs to guarantee expression of the transgene in numerous genetically engineered crop plants (e.g. maize, soy, cotton and canola species)
  - The method will be validated for as many crop events carrying the 35S promoter as possible
- **Customers:** Biotech crop testing laboratories around the world. Broad adoption by world wide testing laboratories would contribute to international harmonization

Some Issues Faced When Developing Reference Materials

- Initial selection of material (SRM components) was for a specific purpose usually and may not address every need in the future (a new locus may not exhibit a diverse set of alleles)
- The forensic community uses commercial STR typing kits – and only wants a confirmation of the allele calls against an allelic ladder – should we fully sequence every sample?
- Some genetic loci will not be able to have every allele sequenced (e.g., due to locus duplication)
- There are lots of loci that could be "certified" – how do we decide which ones to include in future certificate updates?

STRBase: a Community Resource for Forensic DNA Applications of STRs…

A similar standard information resource could be developed for clinical diagnostics or agricultural biotech crop applications

NIST *Applied Genetics*

Thank You for Your Attention…

Contact Information

john.butler@nist.gov

301-975-4049