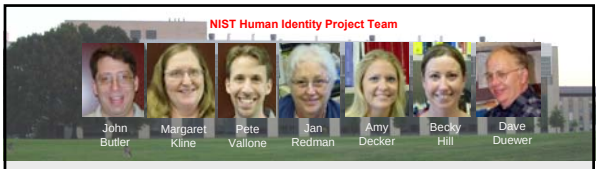


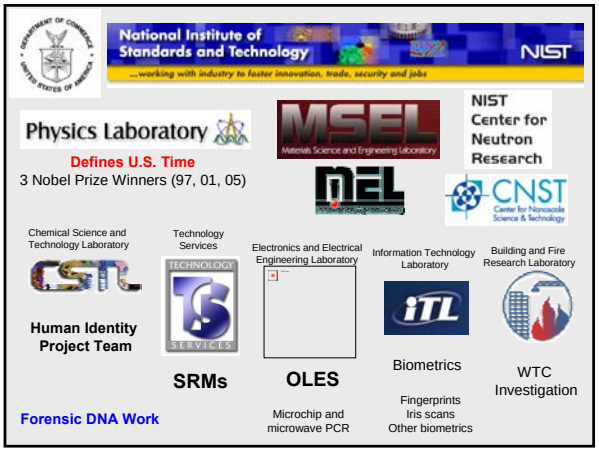
NIST Update

John M. Butler
and Human Identity Project Team
National Institute of Standards and Technology

SWGDAM
Fredericksburg, VA
July 17, 2008

Presentation Outline

- NIST projects overview
- SRM update: SRM 2372, 2391b, 2392, 2395
- STR allele sequencing (Margaret Kline)
- Mixture work (Amy Decker & Michelle Burns)
- Additional STRs and 26plex (Becky Hill)
- Rapid PCR for DNA biometrics (Pete Vallone)
- SNPs and ethnicity estimation (Pete Vallone)
- Training workshops (John Butler/Pete Vallone)



Mark Stolorow is now Director of the NIST Office of Law Enforcement Standards (OLES)

http://www.eeel.nist.gov/oles/directory.html

Sue Ballou
(Forensics)

http://www.eeel.nist.gov/oles/forensics.html

Forensics Research at NIST

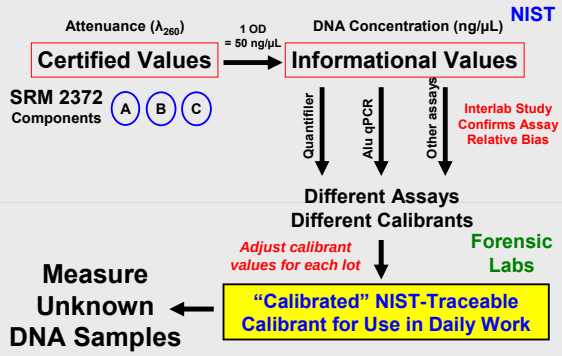
- Computer (digital evidence) forensics
- Ballistics
- Fingerprints for biometrics
- Arson investigation
- DNA

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

Overview of SRM 2372 Values and Use



SRM 2391b and 2395 Certificate Updates

- **SRM 2391b** (Autosomal STR Loci)
  - **MiniFiler examined** (allele dropout with component 8 and D16S539)
  - **Additional Loci: 26 new miniSTR loci**
  - Demonstrating extended stability (new quantitation data and no significant degradation to existing components)

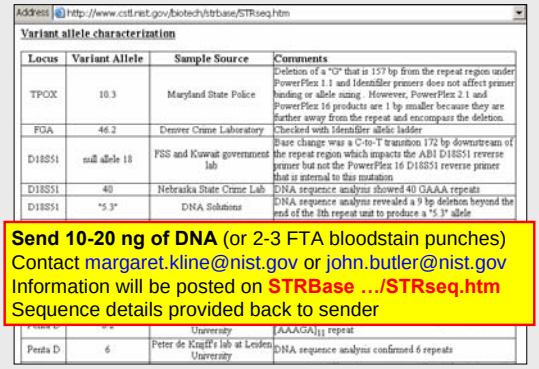
  http://www.cstl.nist.gov/biotech/strbase/srm2391b.htm
- **SRM 2395** (Y-STR and Y-SNP Loci)
  - **Yfiler loci sequenced** (DYS635 now included)
  - **Additional Loci: 20 new Y-STR loci**
  - Demonstrating extended stability (new quantitation data and no significant degradation to existing components)

  http://www.cstl.nist.gov/biotech/strbase/srm2395.htm

Final Documents Submitted, Information Posted on STRBase and Registered Users will be Notified of Certificate Updates

Unusual STR Allele Characterization (Free)

Send us any unusual variant or null alleles and we will sequence them…


Address http://www.cstl.nist.gov/biotech/strbase/STRseq.htm

Variant allele characterization

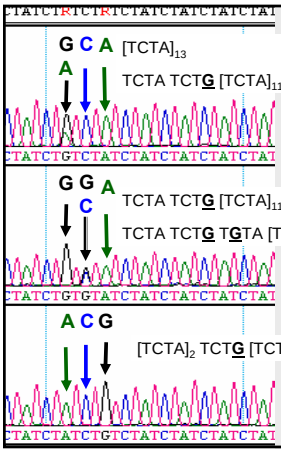
Locus	Variant Allele	Sample Source	Comments
TPOX	10.3	Maryland State Police	Deletion of a "G" that is 157 bp from the repeat region under PowerPlex 1.1 and Identifiler primers does not affect primer binding or allele sizing. However, PowerPlex 2.1 and PowerPlex 16 products are 1 bp smaller because they are further away from the repeat and encompass the deletion.
FGA	46.2	Denver Crime Laboratory	Checked with Identifiler allelic ladder
D18S51	null allele 18	FSS and Kuwait government lab	Base change was a C-to-T transition 172 bp downstream of the repeat region which impacts the ABI D18S51 reverse primer but not the PowerPlex 16 D18S51 reverse primer that is internal to this mutation
D18S51	40	Nebraska State Crime Lab	DNA sequence analysis showed 40 GAAA repeats
D18S51	"5.3"	DNA Solutions	DNA sequence analysis revealed a 9 bp deletion beyond the end of the 8th repeat unit to produce a "5.3" allele
		University	$[AAAGA]_{11}$ repeat
Penta D	6	Peter de Knijff's lab at Leiden University	DNA sequence analysis confirmed 6 repeats

Send 10-20 ng of DNA (or 2-3 FTA bloodstain punches)
Contact margaret.kline@nist.gov or john.butler@nist.gov
Information will be posted on **STRBase …/STRseq.htm**
Sequence details provided back to sender

Locus Sequence Variability

- In collaboration with Tom Hall (IBIS) has analyzed some of the NIST population samples by Mass Spec methods.
- In many samples the Mass Spec detected SNPs, base pair changes within specific loci.
- We have gone back and sequenced some of these samples to determine where the SNPs were located.

SNPs within the D8S1179 repeat



Repeat is TCTA
Three NIST samples have genotypes 13,13.

Analysis by Mass Spec indicates the presence of SNPs (Tom Hall, IBIS)
Confirmation of the Mass Spec by sequencing at NIST indicates:

There are **4** different 13 alleles in these 3 samples.

Base Pair difference between Repeats

D8S1179

[TCTA]$_{13}$
TCTA TCT**G** [TCTA]$_{11}$
TCTA TCT**G** T**G**TA [TCTA]$_{10}$
[TCTA]$_2$ TCT**G** [TCTA]$_{10}$

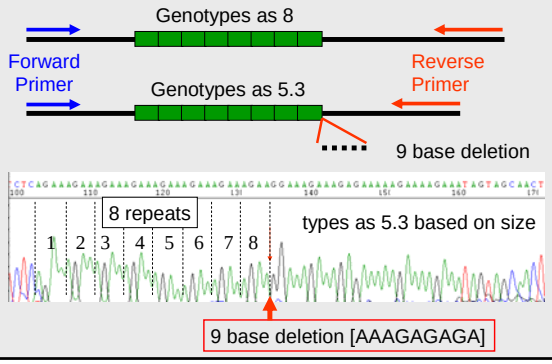
Allele	AVG	SD	N	
8	123.82	0.02	10	
9	127.90	0.02	7	4.1
10	132.03	0.06	42	4.1
11	136.17	0.05	35	4.1
12	140.42	0.07	80	4.3
13	144.93	0.08	153	4.5
14	149.36	0.07	131	4.4
15	153.67	0.07	89	4.3
16	157.93	0.07	21	4.3
17	162.02	0.04	2	4.1

D16S539

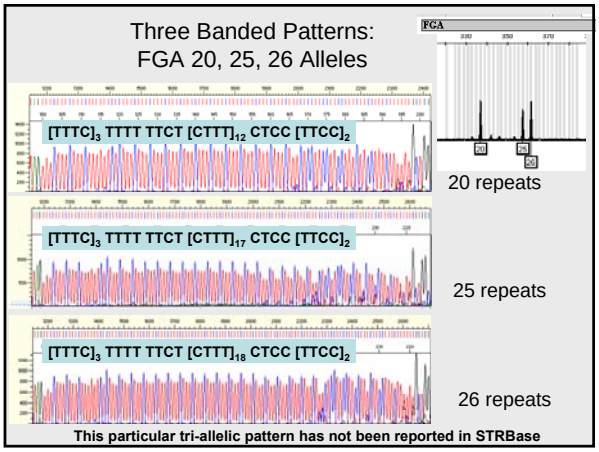
Mass Spec detected fewer SNPs in this locus.
Less average bp variability seen between repeat sizes.

Allele	AVG	SD	N	
5	252.08	0.01	2	
8	264.09	0.11	17	4.0
9	268.14	0.11	93	4.0
10	272.20	0.12	56	4.1
11	276.18	0.14	162	4.0
12	280.25	0.14	158	4.1
13	284.30	0.12	72	4.1
14	288.35	0.07	8	4.1

D18S51 deletion results in 5.3 Allele



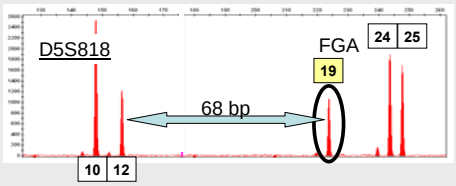
Three Banded Patterns: FGA 20, 25, 26 Alleles



This particular tri-allelic pattern has not been reported in STRBase

Is this an FGA - Tri-allelic pattern identified using Identifiler?

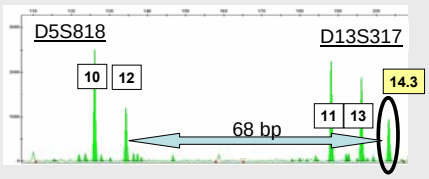
PK HT Ratio
12/10 - 0.48



PK HT Ratio
19/24 - 0.55
25/24 – 0.89

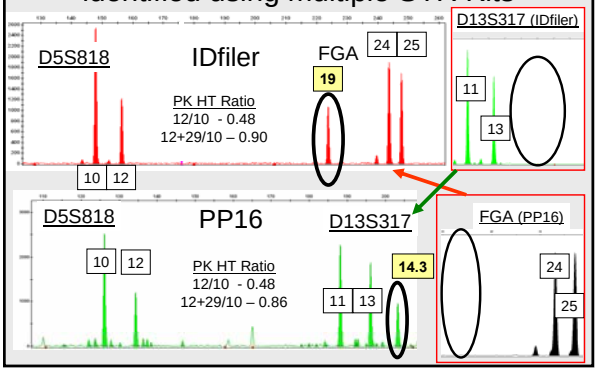
Or is this a D13S317 - Tri-allelic pattern identified using Powerplex 16?

PK HT Ratio
12/10 - 0.48

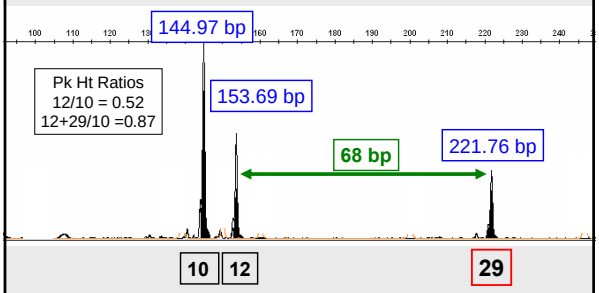


PK HT Ratio
13/11 - 0.83
14.3/11 – 0.42

It's really a D5S818 Tri-allelic pattern identified using multiple STR Kits

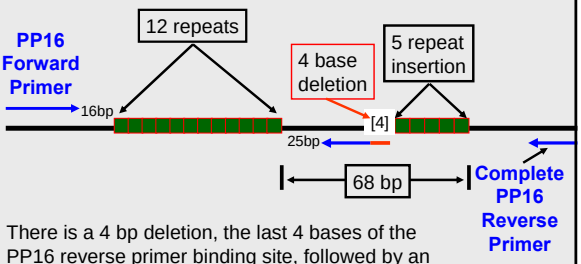


D5S818 monoplex results



The 68 bp size difference between the 12 allele and the variant allele sizing as an **"apparent 29"** allele.

D5S818 Apparent 29 Allele Sequencing Results



There is a 4 bp deletion, the last 4 bases of the PP16 reverse primer binding site, followed by an insertion of 5 repeats.
The 10 and 12 alleles of this sample have been sequenced and have the expected sequences.

Are there other large D5S818 alleles?

- STRBase Tri-allelic reports for FGA for 19,*,* patterns with AB amplification kits.
  - 5 reports :
  - 19,20,21; 19,20,23; 19,20,24; 19,22,23; 19,24,25
  - But there we have sequenced true tri-allelic FGA samples
- STRBase Tri-allelic reports for D13S317 for *,*, OL patterns with PP16 amplification kits.
  - NO tri-allelic patterns with Off-Ladder alleles reported

Mixture Work

- Testing software tools
  - FSS-i3
  - DNA_DataAnalysis (US Army Crime Lab)
- Examining reproducibility of mixture replicates to see how well mixture ratios hold across loci
- Peak height ratio studies with multiple data sets to understand mixture ratio ranges

New Autosomal STR Loci

Aren't the Current STR Loci Good Enough?

- For general forensic matching of evidence to suspect, the 13 CODIS STR loci are sufficient
- For other human identity/relationship testing questions, more autosomal loci can be beneficial or even necessary

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**

New STR Loci Characterized

Hill *et al.* (2008) *J. Forensic Sci.* 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

- Primer sequences, GeneMapper bins and panels, genotypes on common samples, and allele frequency information **available on STRBase**

http://www.cstl.nist.gov/biotech/strbase/miniSTR.htm
http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_NC_loci_types.htm
http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_Panels_Panels.txt
http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_Panels_NC_bins_bins.txt

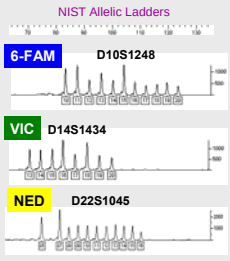
Multiple Miniplexes

- **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

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



NC01 Loci

NC Miniplexes

NC01	NC02	NC03	NC10
D10S1248	D1S1677	D3S3053	D3S3053
D14S1434	D2S441	D6S474	D6S474
D22S1045			D20S482

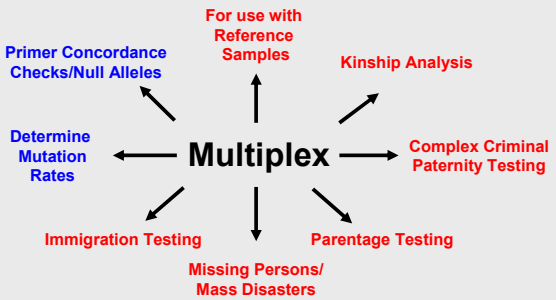
26 total new loci

NC04	NC05	NC06
D1GATA113	D1S1627	D3S4529
D2S1776	D8S1115	D9S2157
D4S2408	D9S324	D10S1430

NC07	NC08	NC09
D9S1112	D17S1301	D10S2327
D12ATA63	D18S8534	D11S4463
D14S1280	D20S1082	D17S974

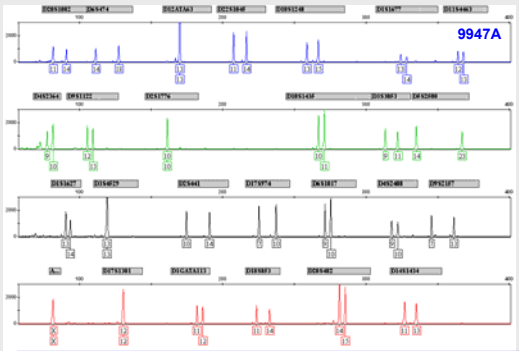
Removed because they were problematic

Why Build a Large Multiplex?



"Autoplex" (26plex)

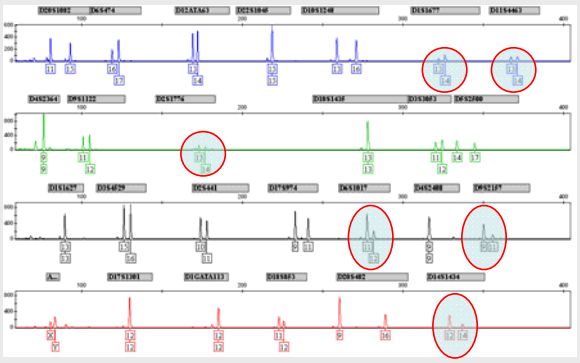
Hill et al. AAFS 2008 talk (Washington, DC) and poster at DNA in Forensics 2008 meeting (Ancona, Italy)



Gender identification + 25 autosomal STR loci in a single amplification

NIST 26plex

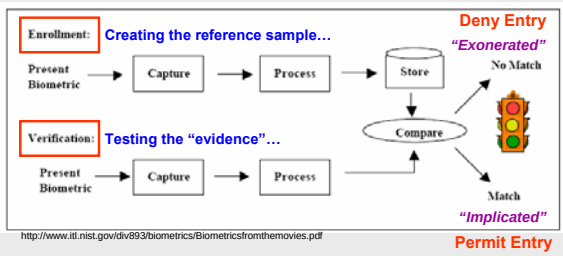
32 Cycles, 50 pg



Heterozygote peak imbalance due to stochastic effects with low DNA template

Rapid PCR
(For DNA Biometrics)

DNA within the Biometric Model



http://www.itl.nist.gov/div893/biometrics/Biometricsfromthemovies.pdf

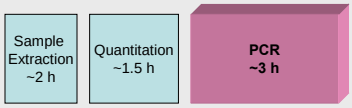
Match of 13 points (each with 2 variable alleles) within DNA

String of 26 numbers (order of listing DNA results would have to be standardized)
16,17-17,18-21,22-12,14-28,30-14,16-12,13-11,14-9,9-11,13-6,6-8,8-10,10

Growing Interest in DNA Biometrics

- DNA sample analysis times are too long (typically 5-6 hours now)
  - Longest part of the process is PCR
  - How fast can we go?
- Other genetic loci could be used (non-CODIS)
  - 26 new miniSTRs developed at NIST

Typical STR DNA Analysis Workflow



Data Review ?

How can we reduce the time needed for cycling?

What happens when we alter cycling parameters?
How will existing commercial kits work?
How will different polymerases perform?
How robust will the results be?
Can we develop novel assays and further the understanding/limits of rapid multiplex PCR?

Rapid PCR Project at NIST

Rapid PCR

- What do we mean by rapid PCR?
  - Rapid hot start polymerases (save ~10min)
  - Shortening cycling hold times (5 sec vs 1 min)
  - Utilizing existing thermal cycling technology (AB 9700)
    - Eliminating 1 ºC/sec ramp rate (9600 emulation)
    - Utilize the 9700 4 ºC/sec ramp rate
  - Using commercial polymerases that are 'faster'

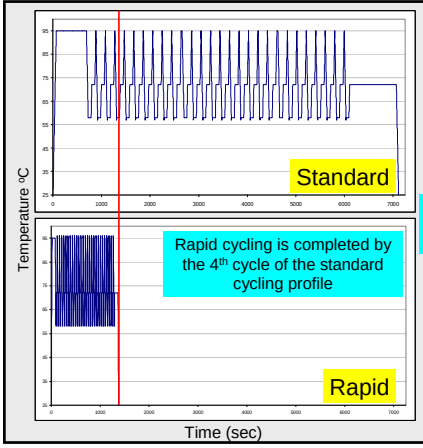
Obtain results in less than 45 minutes
Trying simple things first…

Thermal Cycling

Parameter	Unit	Trad	Rapid	Difference (min)	%
Hot Start	Min	10	1	9.0	6.3
Hold	Sec	60	5/10	72.3	50.6
Soak	Min	60	1	59.0	41.2
Ramp rate	(deg/sec)	1	4	22.4	15.7
Cycles		28	28		
Time		2:58:41	0:35:38	2:23:03	

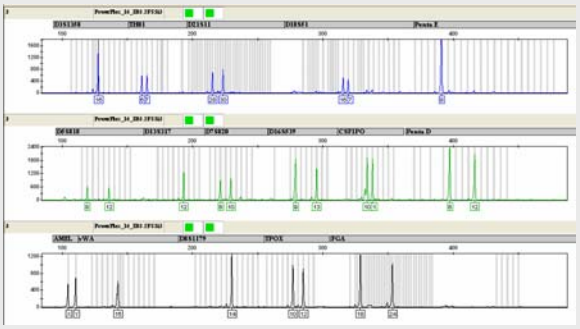
Parameter	Purpose
Hot Start	Primer Dimer, non-specific amplification
Hold	Denature, annealing, elongation, Inter and intra locus balance
Soak	Full adenylation of PCR products

Evaluate robustness and reproducibility

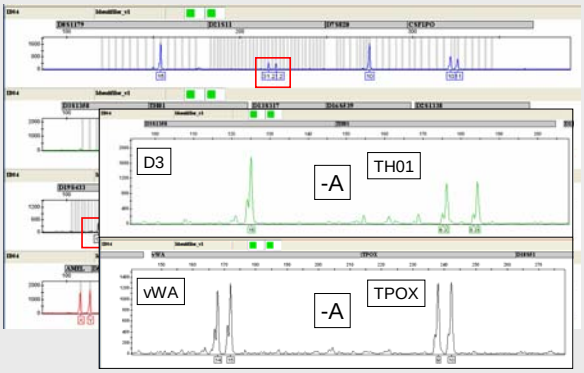


Comparison of Thermal Cycling Profiles On a AB 9700

PowerPlex 16 Rapid Cycling

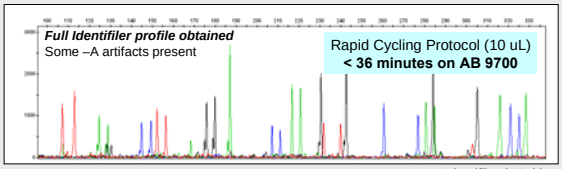


Identifiler Rapid Cycling



Rapid Multiplex PCR Protocols

Testing the potential of rapid multiplex PCR methods
Utilizing AB 9700 cycler and 'fast' commercial enzymes
Manuscript submitted to FSI Genetics in March 2008



Identifiler STR kit
28 cycles, 1ng template DNA

Initial results presented by at 60th Annual Meeting of the American Academy of Forensic Sciences (Washington, DC), February 23, 2008, "Developing Rapid PCR Multiplex Assays with miniSTR Loci"

More recent results on a different thermal cycler in as little as ~15 minutes…

NYC Forensic DNA "X-Prize"

January 17, 2008 Press Release
From Mayor Bloomberg's **STATE OF THE CITY ADDRESS**

"The City will establish **a six-figure prize** for anyone who can invent **a device** tailored to the NYPD **which analyzes the DNA of potential suspects right at the crime scene** - so that officers can release innocent suspects before they are arrested, and track down promising leads more quickly"

http://home2.nyc.gov/html/om/html/2008a/pr017-08.html

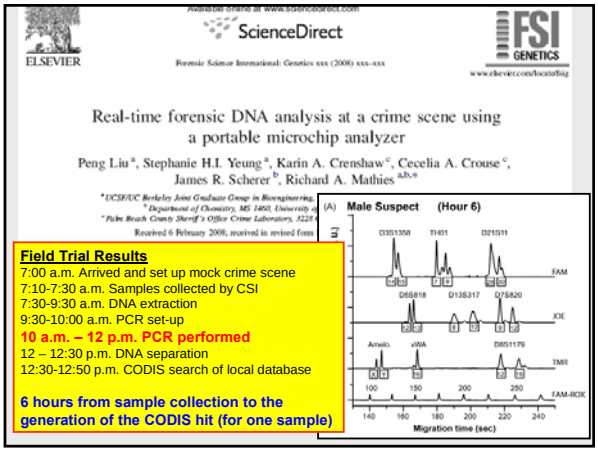
Efforts towards Portable/Mobile DNA Devices

- NEC (Japan)
  - Poster at Promega meeting in Hollywood, CA (Oct 1-4, 2007)
  - Press release on October 15, 2007 (http://www.nec.co.jp/press/en/0710/1501.html)
- Network Biosystems (based on Dan Ehrlich's work at Whitehead)
  - http://www.netbio.com
- Mathies group at UC-Berkeley and Microchip Biotech
  - Publications… in *Analytical Chemistry*, *FSI Genetics*, etc.
  - http://www.microchipbiotech.com
- Landers group at UVA and MicroLab Diagnostics
  - Publications… *Proc Natl Acad Sci USA* 2006; 103:19272-19277
  - http://www.microlabdiagnostics.com

Press Release from NEC (Japan)

NEC Develops World's First Fully Integrated Portable DNA Analyzer
New device to aid criminal investigations and crime prevention through expanded and accelerated use of DNA analysis for individual profiling

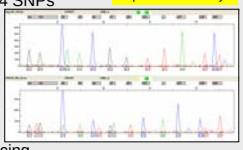
Tokyo, October 15, 2007 - … DNA analysis process consists of 5 steps: (1) cell collection, (2) DNA extraction, (3) Polymerase Chain Reaction (PCR) to amplify DNA fragments, (4) electrophoresis to ascertain DNA "fingerprints" and (5) STR analysis for determining genetic profiling.
…
the device can complete the entire process, from DNA extraction to analysis, in approximately 25 minutes.



Ancestry Informative Markers
SNP Work

- Working with Dr. Manfred Kayser (Netherlands)
  - Set of Ancestry Informative Markers (AIMs)
  - NIST developed assays for typing 24 SNPs
  - Typed over 700 of our samples
- Dr. Peter deKnijff (Netherlands)
  - Performing Y SNP typing
- Dr. Michael Coble (AFDIL)
  - mitochondrial control region sequencing
- How do ancestry markers correlate with self identified ancestry in NIST U.S. samples?

12plex SNP assays



Work presented in Ancona, Italy May 29th 2008
Talk http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm#Presentations

NIST U.S. Sample Set

- 710 NIST samples typed
- African American = 259
- U.S. Caucasian = 262
- U.S. Hispanics = 140
- U.S. Asians = 49

NIST samples have been extensively typed on other forensically relevant markers
Autosomal STRs (CODIS and additional), Y STRs, Y SNPs, Autosomal SNPs, mtDNA sequencing

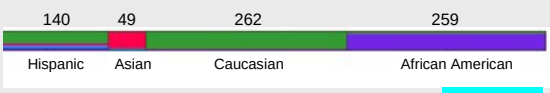
Anonymous

Self Identified Ancestry

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

STRUCTURE Result for 24 SNPs

CEPH

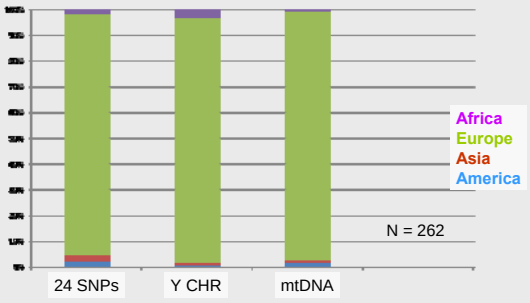


NIST data

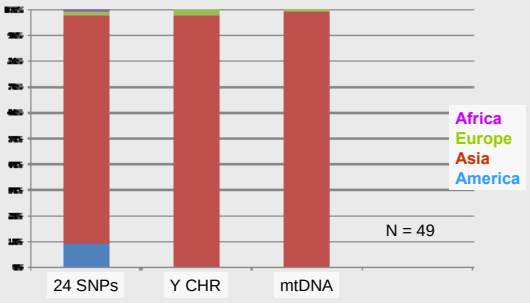
CEPH–Human Genome Diversity Project Cell Line Panel (CEPH-HGDP)

Self ID Ancestry U.S. Caucasian

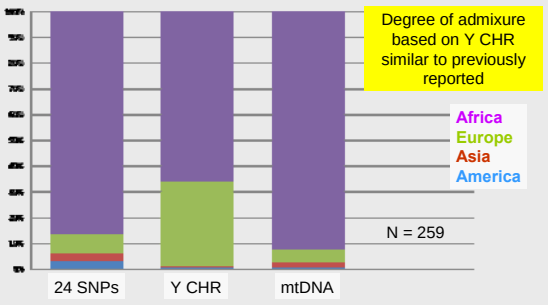
All AIMs correlate well > 90 %



Self ID Ancestry U.S. Asian

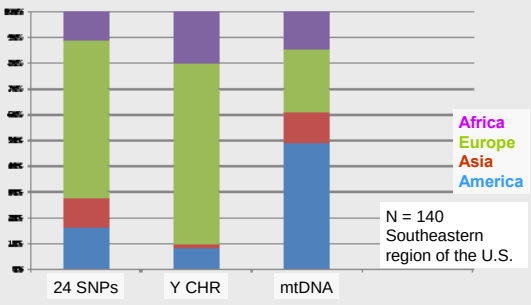


Self ID Ancestry African American



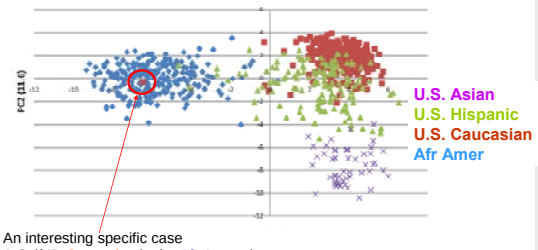
Kayser et al., Genome Res 2003 13:624–34
Vallone and Butler Forensic Sci. 2004 49:723-732

Self ID Ancestry U.S. Hispanic



Further Details in Data Set

PCA based on 24 markers



An interesting specific case
A Self ID Caucasian in the Afr Amer cluster
mtHg L3e2a (Africa)
YHg I (Europe)

qPCR Workshop

- AAFS (February 18th, 2008)
  - **Human DNA Quantification Using Real-Time PCR Assays**
  - **Peter Vallone (NIST)**
  - Margaret Kline (NIST)
  - Eric Buel (Vermont)
  - Jan Nicklas (Vermont)
  - Marie Allen (Uppsala)
  - Mark Timken (CA DOJ)
  - David Foran (Michigan State)
  - Melanie Richard (CFS – Toronto)
  - Toni Diegoli (AFDIL)

158 page handout prepared

http://www.cstl.nist.gov/biotech/strbase/training/AAFS2008_qPCRworkshop.htm

Mixture Interpretation Workshop

- AAFS (February 19, 2008)
  - **DNA Mixture Interpretation:** Principles and Practice in Component Deconvolution and Statistical Analysis
  - **John Butler (NIST)**
  - Ann Gross (MN)
  - George Carmody (Carleton U.)
  - Gary Shutler (WA)
  - Joanne Sgueglia (MA)
  - Angela Dolph (Marshall U./NIST)
  - Tim Kalafut (USACIL)

196 page handout prepared

http://www.cstl.nist.gov/biotech/strbase/training/AAFS2008_MixtureWorkshop.htm

Recent Mixture Training and Discussions

40 analysts from 12 labs

Maryland DNA Analysts

"CE User's Group" April 10, 2008

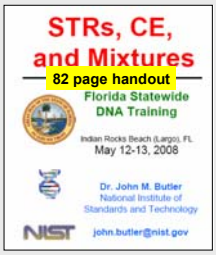
4 hour discussion on mixtures

Have received great feedback on specific needs for guidelines and training – discussed with SWGDAM mixture committee

>80 analysts from 10 labs

Florida Statewide DNA Training

http://www.cstl.nist.gov/biotech/strbase/training/Butler-FL-DNA-Workshop2.pdf



…strbase\training\FL-May2008-Workshop.htm

Planned Promega 2008 Meeting
Troubleshooting Workshop

- Title: **"Principles of Interpretation and Troubleshooting of Forensic DNA Typing Systems"**
- Instructors: **John Butler (NIST) and Bruce McCord (FIU)**
- Date: **October 16, 2008** with Promega Int. Symp. Human ID

The workshop will consist of three parts:
(1) a through examination of theoretical issues with capillary electrophoresis PCR amplification of short tandem repeat markers
(2) a discussion of how to properly set instrument parameters to interpret data (including mixtures), and
(3) **a review of specific problems seen by labs** submitting problematic data and commentary on possible troubleshooting solutions.

Seeking input of problems observed with CE systems

Thank you for your attention…

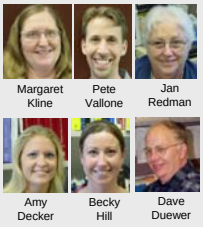
Our team publications and presentations are available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

Questions?

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

john.butler@nist.gov

301-975-4049



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

Collaborators (SNP work)
Manfred Kayser
Oscar Lao
Mike Coble
Jodi Irwin
Peter de Knjiff
Jeroen Pijpe
Kristiaan van der Gaag

Disclaimers

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

Points of view are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice.

Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

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