

SWGDAM (Dumfries, VA) January 17, 2013

# **NIST Update**

### **Michael D. Coble**

### **NIST Applied Genetics Group**

National Institute of Standards and Technology Gaithersburg, Maryland





### **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

#### **DNA Biometrics Team**

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





Coble

mtDNA & Y

John Butler

Mike

Becky Hill

STRBase. Workshops Mixtures. & Textbooks

Concordance & LT-DNA SRM work, variant alleles & Cell Line ID



Kline

Data Analysis Support

Dave

Duewer



Pete Vallone



**Butts** 



Kevin **Kiesler** 

Rapid PCR. Direct PCR & Biometrics

ABI 3500 & DNA Extraction

PLEX-ID & NGS Exploration







**Office Manager** Patti Rohmiller

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



# **APPLIED GENETICS Group**

Major Programs Currently Underway

### Forensic DNA

- STRBase website
- New loci and assays (26plex)
- STR kit concordance
- Ancestry SNP assays
- Low-template DNA studies
- Mixture interpretation research and training
- STR nomenclature
- Variant allele cataloging and sequencing
- ABI 3500 validation
- Training workshops to forensic DNA laboratories
- Validation experiments, information and software tools
- Textbooks 3<sup>rd</sup> ed. (3 volumes)

Clinical Genetics

- Huntington's Disease SRM
- CMV SRM
- Exploring future needs
- **DNA Biometrics** 
  - Rapid PCR methods
  - Testing of rapid DNA systems
  - Plex-ID mtDNA base composition

### Cell Line Authentication

 ATCC documentary standard (Margaret Kline & John Butler served on this international committee)

### **Aiding Cell Line Authentication**

Katsnelson, A. (2010) Nature News, 465: 537 (3 June 2010)

# **Biologists tackle cells' identity crisis**

DNA fingerprinting scheme aims to make sure researchers are working on the right cells.

Ever since biologists learned how to grow human cells in culture half a century ago, the cells have been plagued by a problem of identity: many commonly used cell lines are not actually what researchers think they are.

Cell-line misidentification has led to mistakes in the literature, misguided research based on those results and millions wasted in grant money. Last year, *Nature* described the situation as a scandal<sup>1</sup>.

But a universal system for determining the identity of cell lines may now be in view. Next month, a working group led by the American Type Culture Collection (ATCC), a nonprofit biological repository based in Manassas,

Virginia, that stores 3,600 cell lines from more than 150 species, plans to unveil standard-



ATCC<sup>®</sup> Standards Development Organization

#### **Designation: ASN-0002**

#### Authentication of Human Cell Lines: Standardization of STR Profiling

The working group, composed of representatives from academia, government and industry, a universally accepted approach will allow different facilities to compare their cell lines with each other, he adds.

Fingerprinting has its limits, cautions Michael Johnson, a cancer researcher at Georgetown University in Washington DC. "Just because a cell fingerprints out as the same [as another cell] doesn't mean they will behave the same," he says, noting that a cell's properties can also be affected by the way it has been grown, the number of times it has been cultured anew and small genetic changes that wouldn't show up in a fingerprint test. One classic example, he notes, is an immortalized breast cell line called MCF10A, which can form organized hollow

structures similar to those found in mammary tissue; MCF10A cells currently distributed by

http://www.nature.com/news/2010/100602/pdf/465537a.pdf

# Highlights Since Last SWGDAM

- InDel work published
- PLEX-ID report available
- New DNA mixture training materials
- TrueAllele evaluation continues...
- New autosomal STR and Y-STR loci & kits
   NIST U.S. population data set completed
- SRM 2372 recertified
- Rapid DNA efforts
- Interpretation book being written

# Insertion/Deletion (InDel) Markers



lituto de Patologia e Imunologia Molecular da Universidade do Porto





Manuel Fondevila Alvarez Guest Researcher from Spain (Jan 2011 to July 2012)



#### Main Points:

 InDels (insertion-deletion) or DIPs (deletion-insertion polymorphisms) are short length polymorphisms, consisting of the presence or absence of a short (typically 1-50 bp) sequence

NIST

- Like SNPs, InDels have low mutation rate (value to kinship analysis), small amplicon target sizes (value with degraded DNA), and can be highly multiplexed
- Can be analyzed on CE instruments like STRs
- Studied commercial 30plex (Qiagen DIPlex) and a home-brew 38plex in U.S. population samples

Int J Legal Med (2012) 126:725-737 DOI 10.1007/s00414-012-0721-7

Int. J. Legal Med. (2012) 126: 725-737

ORIGINAL ARTICLE

Forensic performance of two insertion-deletion marker assays

M. Fondevila • C. Phillips • C. Santos • R. Pereira • L. Gusmão • A. Carracedo • J. M. Butler • M. V. Lareu • P. M. Vallone

# Performance Assessment of Plex-ID

Abbott Ibis Biosciences Plex-ID System



**Kevin Kiesler** 



NIST Report to the FBI: Plex-ID Electrospray Time-of-Flight Mass Spectrometer for Mitochondrial DNA Base Composition Profiling

Experiments performed and report written by: Kevin Klesler, M.S. (NIST)

Under the direction of: Dr. Peter Vallone (NIST)

- In collaboration with FBI
- Evaluating ESI-TOF mass spectrometer for mtDNA
- Base composition of the control region determined from 8 triplex PCRs
- Started running the Plex-ID platform mid-October 2011
  - 136 page NIST report available on STRBase

http://www.cstl.nist.gov/strbase/pub\_pres/NIST-report-on-PlexID.pdf

# Mixture Training Workshops



### **MIXTURE INTERPRETATION WORKSHOP**

### Mixtures Using SOUND Statistics, Interpretation & Conclusions

23<sup>rd</sup> International Symposium on Human Identification October 15, 2012 (Nashville, TN)

#### Presenters

John M. Butler, PhD Michael D. Coble, PhD Robin W. Cotton, PhD Catherine M. Grgicak, PhD Charlotte J. Word, PhD NIST, Applied Genetics Group NIST, Applied Genetics Group Boston University, Biomedical Forensic Sciences Boston University, Biomedical Forensic Sciences Consultant



John Butler Mike Coble

- Collaborators from Boston University (formerly Cellmark)
- ISHI 2012 workshop covered issues with thresholds, statistics, probabilistic genotyping, complex mixtures, court testimony, and assumptions made
  - Audience response systems (clickers) used to gather data from participants
- Slides are available on STRBase

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

# December 2012 Issue of FSI Genetics



Editorial

Focus issue—Analysis and biostatistical interpretation of complex and low template DNA samples



DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods

P. Gill<sup>a,b,\*</sup>, L. Gusmão<sup>c</sup>, H. Haned<sup>d</sup>, W.R. Mayr<sup>e</sup>, N. Morling<sup>f</sup>, W. Parson<sup>g</sup>, L. Prieto<sup>h</sup>, M. Prinz<sup>i</sup>, H. Schneider<sup>j</sup>, P.M. Schneider<sup>k</sup>, B.S. Weir<sup>1</sup>

### Some of the articles present in this issue...



### **TrueAllele Mixture Software Evaluation**

### Main Points:

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

### **Presentations/Publications**:

- ISFG 2011 presentation
- Numerous mixture workshop talks (see http://www.cstl.nist.gov/strbase/mixture.htm)

#### D19S433 result from one replicate of 50,000 simulations





# STR Kit Concordance Studies



**Becky Hill** 



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



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

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

- Examined NIST samples across >20 STR kits and inhouse 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 Oct 2012)
- Information provided back to kit developers to redesign primers or add extra ones – often prior to kit release

### Aiding Improvements with SE33 Primers



Robert S. McLaren<sup>1</sup>, Jaynish Patel<sup>1</sup>, Douglas R. Storts<sup>1</sup>, Carolyn R. Hill<sup>2\*</sup>, Margaret C. Kline<sup>2</sup> and John M. Butler<sup>2</sup>

<sup>1</sup>Promega Corporation

<sup>2</sup>Human Identity Project Team, National Institutes of Standards and Technology

Publication Date: 2012

#### A developmental validation article is in press

### PowerPlex ESI 17 Pro vs ESI 17 SE33 Results



### PowerPlex ESI 17 Pro SE33 allele 23.2

Reverse primer is inside of hairpin region

The SE33 locus range is shown for both PowerPlex® ESI 17 Pro (Panel A) and ESI 17 (Panel B) amplifications of DNA sample GT37190. Peak labels show allele calls (top) and sizes in bases (bottom). The off-ladder peak seen with PowerPlex® ESI 17 is correctly called as 23.2 with the PowerPlex® ESI 17 Pro System

http://www.promega.com/resources/articles/profiles-in-dna/2012/improved-primer-pair-for-the-se33-locus-in-the-powerplex-esi-17-pro-system/

# Variant STR Allele Sequencing

#### Main Points:



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



Short communication

This year we successfully navigated lawyers and legal agreements on both sides to create an MOU with an SDIS lab permitting NIST to



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

Margaret C. Kline<sup>\*</sup>, Carolyn R. Hill, Amy E. Decker<sup>1</sup>, John M. Butler National Institute of Standards and Technology, 100 Bureau Drive, M/S 8312, Gaithersburg, MD 20899, USA

#### Presentations/Publications:

FSI Genetics article (Aug 2011) and numerous talks

# 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

# 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

# **Characterizing New STR Loci**

#### Main Points:





John Butler

**Becky Hill** 

- 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

#### **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

	Alleles	Genotypes	Het	P <sub>I</sub> Value	
Locus	Observed	Observed	(obs)	n=1036	
SE33	52	304	0.9353	0.0066	
Penta E	23	138	0.8996	0.0147	
D2S1338	13	68	0.8793	0.0220	
D1S1656	15	93	0.8890	0.0224	
D18S51	22	93	0.8687	0.0258	
D12S391	24	113	0.8813	0.0271	
FGA	27	96	0.8745	0.0308	
D6S1043	27	109	0.8494	0.0321	
Penta D	16	74	0.8552	0.0382	
D21S11	27	86	0.8330	0.0403	
D8S1179	11	46	0.7992	0.0558	
D19S433	16	78	0.8118	0.0559	
vWA	11	39	0.8060	0.0611	
F13A01	16	56	0.7809	0.0678	
D7S820	11	32	0.7944	0.0726	
D16S539	9	28	0.7761	0.0749	
D13S317	8	29	0.7674	0.0765	
TH01	8	24	0.7471	0.0766	
Penta C	12	49	0.7732	0.0769	
D2S441	15	43	0.7828	0.0841	
D10S1248	12	39	0.7819	0.0845	
D3S1358	11	30	0.7519	0.0915	
D22S1045	11	44	0.7606	0.0921	
F13B	7	20	0.6911	0.0973	
CSF1PO	9	31	0.7558	0.1054	
D5S818	9	34	0.7297	0.1104	
FESFPS	12	36	0.7230	0.1128	
LPL	9	27	0.7027	0.1336	
ΤΡΟΧ	9	28	0.6902	0.1358	

Rank Order of 29 Autosomal STR Loci in Commercial Kits with NIST 1036 U.S. Population Samples

http://www.promega.com/resources/ articles/profiles-in-dna/2012/ variability-of-new-str-loci-and-kits-inus-population-groups/

Hill et al. ISHI 2012 poster #84 (see STRBase); Butler et al. (2012) Profiles in DNA

### Probability of Identity Values for Various STR Kits or Locus Combinations based on NIST 1036 U.S. Population Samples

STR Kit or Core Set of Loci	Total N=1036	Caucasians (n=361)	African Am. (n=342)	Hispanics (n=236)	Asians (n=97)
CODIS 13	5.02E-16	2.97E-15	1.14E-15	1.36E-15	1.71E-14
Identifiler	6.18E-19	6.87E-18	1.04E-18	2.73E-18	5.31E-17
PowerPlex 16	2.82E-19	4.24E-18	6.09E-19	1.26E-18	2.55E-17
PowerPlex 18D	3.47E-22	9.82E-21	5.60E-22	2.54E-21	7.92E-20
ESS 12	3.04E-16	9.66E-16	9.25E-16	2.60E-15	3.42E-14
ESI 16 / ESX 16 / NGM	2.80E-20	2.20E-19	6.23E-20	4.03E-19	9.83E-18
ESI 17 / ESX 17 / NGM SElect	1.85E-22	1.74E-21	6.71E-22	3.97E-21	1.87E-19
CODIS 20	9.35E-24	7.32E-23	6.12E-23	8.43E-23	4.22E-21
GlobalFiler	7.73E-28	1.30E-26	3.20E-27	2.27E-26	1.81E-24
PowerPlex Fusion	6.58E-29	2.35E-27	1.59E-28	2.12E-27	1.42E-25
All 29 autosomal STRs	2.24E-37	7.36E-35	3.16E-37	2.93E-35	4.02E-32
29 autoSTRs + DYS391	1.07E-37	3.26E-35	1.77E-37	1.29E-35	2.81E-32

Hill et al. ISHI 2012 poster #84 (see STRBase); Butler et al. (2012) Profiles in DNA

### STR Kit Layouts by Dye Label and PCR Product Size



### DNA Mixture with **PowerPlex Fusion** (Promega)

#### 24plex assay



# Rapidly Mutating Y-STR Loci



Mike Coble Becky Hill





- Part of RM Y-STR Study Group organized by Manfred Kayser (Erasmus University, The Netherlands)
- Supplied data from 1,296 U.S. samples (634 population + 331 father/son pairs)
- Publication with RM Y-STR Study Group is forthcoming

K. Ballantyne et al. 2010; K. Ballantyne et al. 2012

# Rapidly Mutating (RM) Y-STRs

NIST supplied data from 1,296 U.S. samples (634 population + 331 father/son pairs) to RM Y-STR Study Group led by Manfred Kayser (11,978 samples from 169 worldwide populations)



# PowerPlex Y23 Kit



#### 125pg male + 400ng female (3200x female)



Kit found to be *sensitive* and *specific* to male DNA

- Typed 1032 males
  from 4 U.S.
  population groups
- Data supplied to YHRD and USYSTR databases
- Publications are forthcoming
- Full dataset to be released on STRBase

N = 1032 males	PowerPlex Y	Yfiler	PowerPlex Y23
# haplotypes	891	1013	1029
discrimination capacity	0.863	0.982	0.997
# times haplotype observed	PPY (12 loci)	Yfiler (17 loci)	PPY23 (23 loci)
1	821	998	1026
2	41	12	3
3	16	2	
4	6	1	
5	2		
6	2		
7	1		
8	•		
9	1		
10			
11			
12			
13			
14			
15			
16	•		
17	•		
18	•		•
19	1	_	_

Number of unique and shared haplotypes observed with various combinations of Y-STR loci across 1032 U.S. population samples

1026 PPY23 haplotypes occur once; and3 sets of sample pairs cannot be resolved from one another

### NIST Reference Materials for Forensic DNA Measurement Assurance







**SRM 2372 has been recertified** because the dsDNA has unraveled, which impacts absorbance certification values. We are recertifying the samples with aid of digital PCR measurements. **Now available again!** 

DNA quantity measurement calibration



SRM 2391c currently does not cover the six additional Y-STR markers in PowerPlex Y23. We plan to certify values for these markers by mid-2013.



Autosomal and Y-chromosome short tandem repeat (STR) measurement calibration

# NIST Reference Materials for Forensic DNA Measurement Assurance



ack on the market as of January 8<sup>th</sup>, 2013

March 2012: SRM 2372 was taken off of the market after stability measurement of the material indicated all components had increased in UV absorbance

### Why did it change?

The original tightly coiled double-strands have unraveled, which impacts the original absorbance certification values.

### **Solution for recertification:**

For recertification with absorbance, each of the components was transformed to a **single-stranded confirmation** with the **addition of sodium hydroxide (NaOH)**. INTERNATIONAL ISO STANDARD 21571

There has been no change in the behavior of SRM 2372 for qPCR.

# ABI 3500 Validation Studies

### Main Points:

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

#### **Presentations/Publications**:

- MAAFS talk (May 2011)
- ABI road show talks (July & Aug 2011)
- ISFG presentation (Sept 2011)
- Forensic News (Spring 2012)

### **HID in Action**

3500 Genetic Analyzer: Validation Studies

Erica L.R. Butts and Peter M. Vallone National Institute of Standards and Technology



#### Erica Butts

# Rapid DNA Efforts

Accelerated Nuclear DNA Equipment (ANDE) developed by **NetBlo** 



#### RapidHIT 200 developed by IntegenX





Pete Vallone Erica Butts

- Evaluating ANDE (NetBio) and IntegenX rapid DNA instruments
  - both instruments are capable of swab in → STR profile out in less than 90 minutes without user intervention
- Exploring rapid DNA techniques including direct PCR and rapid PCR
  - STR profiles generated in <2 hours with standard lab equipment and rapid protocols
  - See ISHI 2012 poster available on STRBase "<u>Rapid DNA Testing</u> <u>Approaches for Reference Samples</u>"

Fastest results swab-to-profile (Identifiler): 57 minutes

# Forensic DNA Typing Textbook 3<sup>rd</sup> Edition is Three Volumes

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



John Butler

For beginning students, general public, & lawyers



### Fall 2009

~500 pages



Fall 2011 ~700 pages

Currently being written Advanced Topics in FORENSIC **DNA TYPING: INTERPRETATION** John M. Butler CAG GOC Ala Gln Gly

Fall 2013 ~500 pages

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

Chapter	<b>Topic</b> (current planned chapters)
	Introduction
1	Data interpretation overview
2	Thresholds
3	STR alleles & artifacts
4	STR genotypes & dropout
5	STR profiles
6	Mixture interpretation
7	Low-level DNA and complex mixtures
8	CE troubleshooting
9	Statistical interpretation overview
10	STR population data analysis
11	Profile frequency estimates
12	Mixture statistics
13	Coping with potential missing alleles
14	Kinship and parentage analysis
15	Lineage marker statistics
16	Drawing conclusions & report writing
App 1	U.S. Population Data (29 loci with N=1036)
App 2	NRC I and II Recommendations (1992/1996)
Арр З	DAB Recommendations on Stats (Feb 2000)
App 4	Glossary
App 5	Worked Example for Mixture Interpretation

### **Features in New Book**

(planned for Fall 2013 release)

 Numerous D.N.A. Boxes (Data, Notes, & Applications)

- Worked examples to show relevance of equations
- "Better know a statistician"
- Interviews on report writing from multiple perspectives
- Explanations of SWGDAM interpretation guidelines
- Mixture interpretation
- Kinship analysis
- CE troubleshooting
- Standard U.S. pop data

# Thank you for your attention

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

### **Contact Information**

John Butler NIST Fellow Group Leader john.butler@nist.gov 301-975-4049 Mike Coble Forensic Biologist Applied Genetics Group michael.coble@nist.gov 301-975-4330



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

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