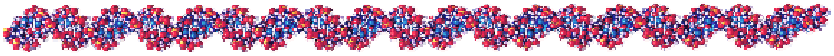
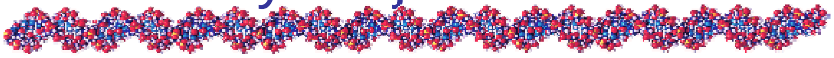


Overview of the Human Identity Projects at NIST



Dr. Peter M. Vallone
Applied Genetics Group
U.S. National Institute of Standards and Technology (NIST)

Presented to FSSA
Adelaide, South Australia
September 13, 2010

What is NIST?

- U.S. National Institute of Standards and Technology
- Founded 1901

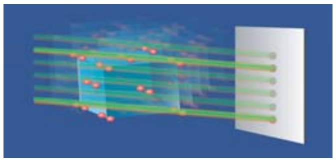
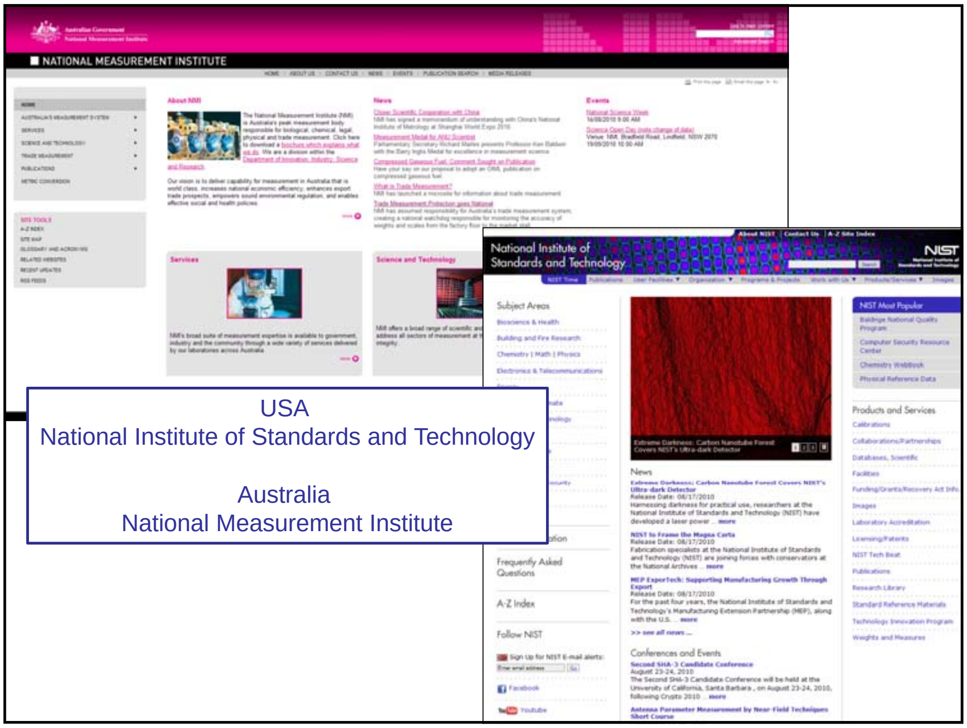


Illustration by Jeffrey Aarons/NIST

- Main campus in Gaithersburg, Maryland
  30 miles north of Washington D.C.

To promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve quality of life.



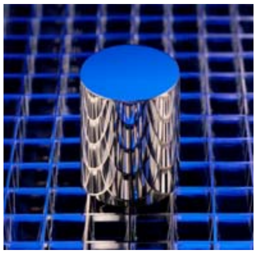
USA
National Institute of Standards and Technology

Australia
National Measurement Institute

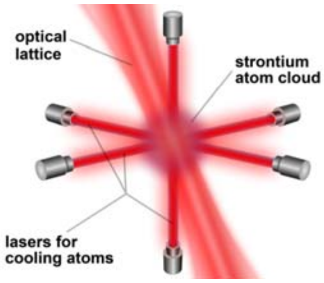
NIST Products and Services

Measurement Research
- ~2,200 publications per year

Standard Reference Data
- ~100 different types
- ~6,000 units sold per year
- ~130 million data downloads per year



Copyright Robert Rathe

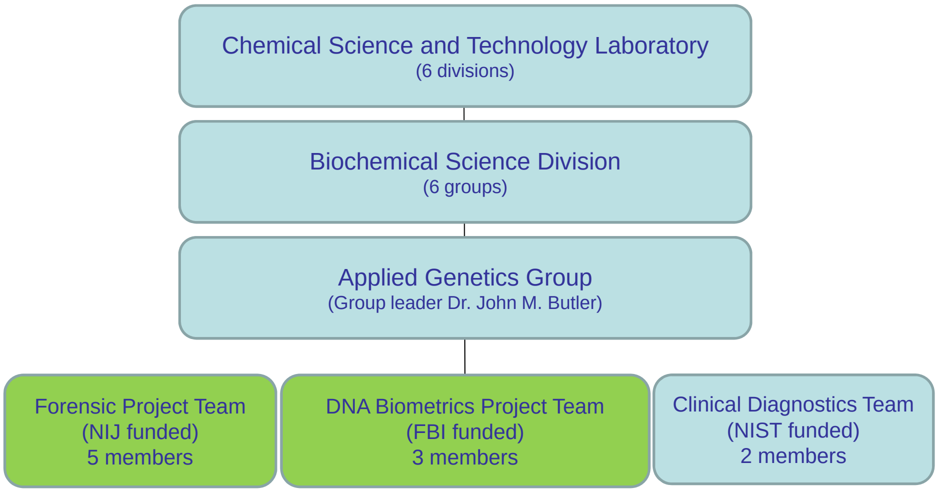


Standard Reference Materials
- ~1,300 products available
- ~33,000 units sold per year

Calibration Tests
- ~16,000 test per year

Laboratory Accreditation
- ~800 accreditations of testing and calibration laboratories per year

NIST Organization



Chemical Science and Technology Laboratory
(6 divisions)

Biochemical Science Division
(6 groups)

Applied Genetics Group
(Group leader Dr. John M. Butler)

Forensic Project Team
(NIJ funded)
5 members

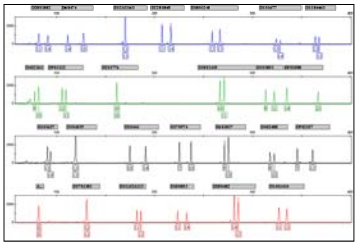
DNA Biometrics Project Team
(FBI funded)
3 members

Clinical Diagnostics Team
(NIST funded)
2 members

Our Group Goals

Provide support for the forensic DNA typing community

- Standards
- Technology
- Training and education



Short Tandem Repeat DNA
Internet DataBase

NIST Standard Reference Database SRD 130

Created by John M. Butler

General Information

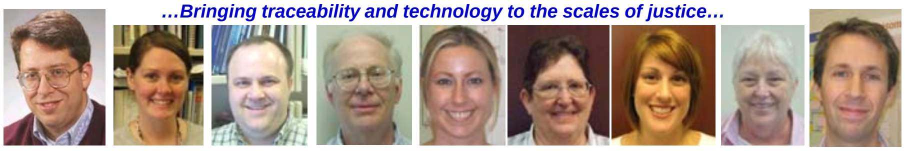
Forensic STR Information

The NIST Human Identity Project Team

(Forensic DNA & DNA Biometrics)

Funding from the **National Institute of Justice (NIJ)** through the NIST Office of Law Enforcement Standards and the **FBI S&T Branch** through the NIST Information Access Division

…Bringing traceability and technology to the scales of justice…



John Butler *Project Leader, Forensic DNA*
Erica Butts
Mike Coble
Dave Duewer
Becky Hill
Margaret Kline
Kristen O'Connor
Jan Redman
Pete Vallone *Project Leader, DNA Biometrics*

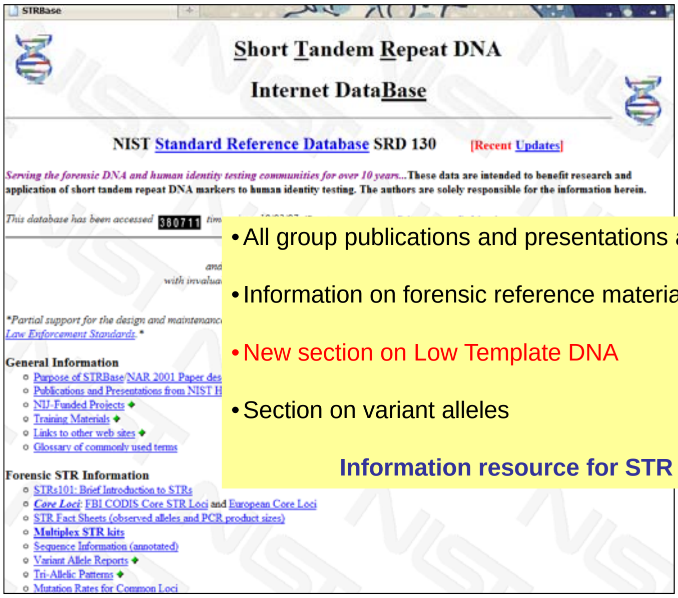
Workshops & Textbooks
Mixtures, mtDNA & Y
Concordance & LT-DNA
Kinship Analysis
Rapid PCR & Biometrics

DNA Extraction Efficiency
Software Tools & Data Analysis
Variant alleles & Cell Line ID
STRBase Support

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

STRBase

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



STRBase

Short Tandem Repeat DNA

Internet DataBase

NIST Standard Reference Database SRD 130 [Recent Updates]

Serving the forensic DNA and human identity testing communities for over 10 years…**These data are intended to benefit research and application of short tandem repeat DNA markers to human identity testing. The authors are solely responsible for the information herein.**

This database has been accessed 380711 *tim*

and
with invalua

**Partial support for the design and maintenanc*
Law Enforcement Standards. *

General Information
- Purpose of STRBase/NAR 2001 Paper des
- Publications and Presentations from NIST H
- NIJ-Funded Projects
- Training Materials
- Links to other web sites
- Glossary of commonly used terms

Forensic STR Information
- STRs101: Brief Introduction to STRs
- *Core Loci*: FBI CODIS Core STR Loci and European Core Loci
- STR Fact Sheets (observed alleles and PCR product sizes)
- **Multiplex STR kits**
- Sequence Information (annotated)
- Variant Allele Reports
- Tri-Allelic Patterns
- Mutation Rates for Common Loci

John Butler

- All group publications and presentations are available
- Information on forensic reference materials
- New section on Low Template DNA
- Section on variant alleles

Information resource for STR typing

Projects within the Forensic and Biometric Project Teams

- Concordance testing with STR kits
- Additional STR loci
- DNA extraction efficiency
- Kinship analysis
- Rapid PCR protocols
- Mixture interpretation
- Low template PCR
- Workshops and training

Concordance Testing


Becky Hill

- When new STR typing kits are released, concordance testing is performed
  - Buffer and master mix improvements
  - PCR primer sequences may change
- Type a set of NIST U.S. population samples (n > 600) to evaluate concordance
- Provides a service to the forensic community and kit manufactures

STR Loci Present in Commercial Kits

U.S.			Europe			
PP16	Identifiler	MiniFiler	ESX/ESI17	NGM	SEfiler	SGM Plus
TPOX	TPOX					
CSF1PO	CSF1PO	CSF1PO				
D5S818	D5S818					
D7S820	D7S820	D7S820				
D13S317	D13S317	D13S317				
FGA	FGA	FGA	FGA	FGA	FGA	FGA
vWA	vWA		vWA	vWA	vWA	vWA
D3S1358	D3S1358		D3S1358	D3S1358	D3S1358	D3S1358
D8S1179	D8S1179		D8S1179	D8S1179	D8S1179	D8S1179
D18S51	D18S51	D18S51	D18S51	D18S51	D18S51	D18S51
D21S11	D21S11	D21S11	D21S11	D21S11	D21S11	D21S11
TH01	TH01		TH01	TH01	TH01	TH01
D16S539	D16S539	D16S539	D16S539	D16S539	D16S539	D16S539
	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338
	D19S433		D19S433	D19S433	D19S433	D19S433
			D12S391	D12S391		
			D1S1656	D1S1656		
			D2S441	D2S441		
			D10S1248	D10S1248		
			D22S1045	D22S1045		
			SE33		SE33	
Penta D						
Penta E						

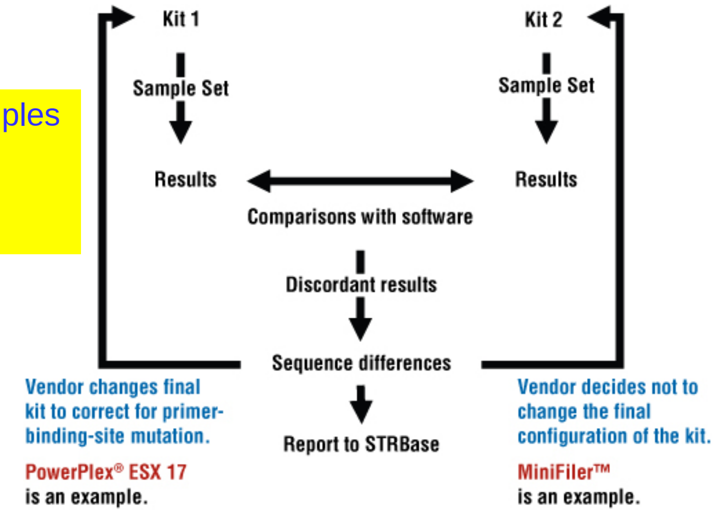
U.S. is looking to expand the core loci (18-20 total)

to provide more international overlap

Strategies for Concordance Testing

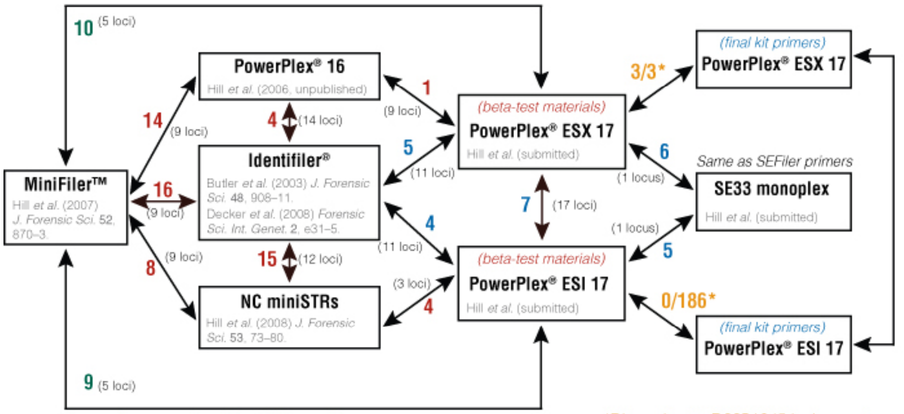
the four S's of concordance studies

Standard samples
Software
Sequencing
STRBase



Profiles in DNA Article Published April 2010 http://www.promega.com/profiles/1301/1301_08.html

Concordance Comparisons



Comparisons made with ~660 samples
Comparisons made with ~1,120 samples
Comparisons made with ~1,440 samples

*Discordant at D22S1045 in three out of three "G to T samples" tested following correction in final kit primers.

**Zero discordant samples out of 186 retested with final kit primers.

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

Additional STR Loci

Pete Vallone

Becky Hill

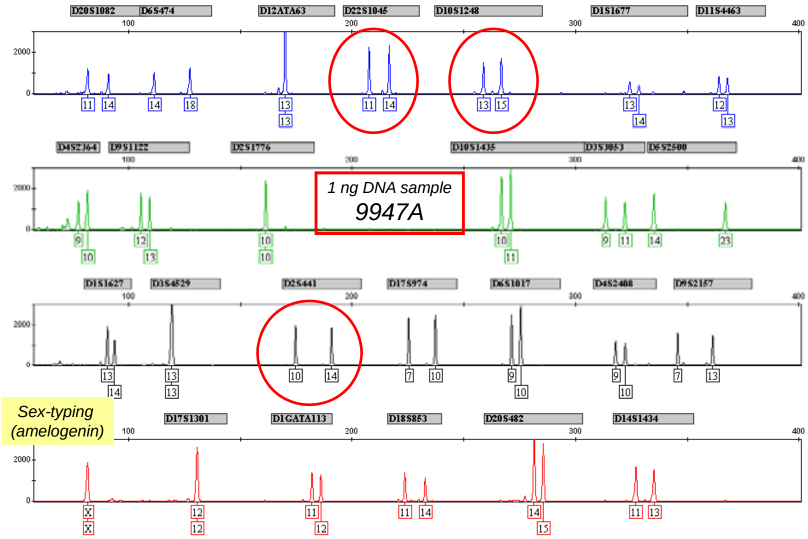
- Characterizing additional STR loci for human identity testing (26 total)
  - Degraded DNA (miniSTRs)
  - Kinship analysis
  - Potential candidates for core loci
  - New core loci adopted by Europe
- In house assays developed for typing
- Population data collected on NIST samples

Europe adopts new loci D10S1248, D14S1434 and D22S1045 Gill *et al. Forensic Sci. Int.* 2006;156:242-244
D2S441 replaces D14S1434 Gill *et al. Forensic Sci. Int.* 2006;163:155-157

Hill, C.R., Kline, M.C., Coble, M.D., Butler, J.M. (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *J. Forensic Sci.* 53(1):73-80.

NIST 26plex published in *J. Forensic Sci.* (Sept 2009)

25 new STR loci (unlinked from 13 CODIS core loci)



Gender identification + 25 autosomal STRs in a single amplification

1 ng DNA sample
9947A

Sex-typing (amelogenin)

Hill, C.R., et al. (2009) A new STR 26plex assay for human identity testing. *J. Forensic Sci.* 54: 1008-1015

Primer sequences and GM bins & panels available: **http://www.cstl.nist.gov/biotech/strbase/str26plex.htm**

DNA Extraction Efficiency

Erica Butts

- Using known amounts of DNA to estimate true extraction efficiency

Sources of DNA:

- Highly characterized extracted DNA (52.44 ng/µL)
- Human cell lines (6 pg of DNA per cell)
- Whole blood (white blood cell count of 4.0 million WBC/mL)

The starting amount of DNA is known

Qiagen EZ1 Advanced

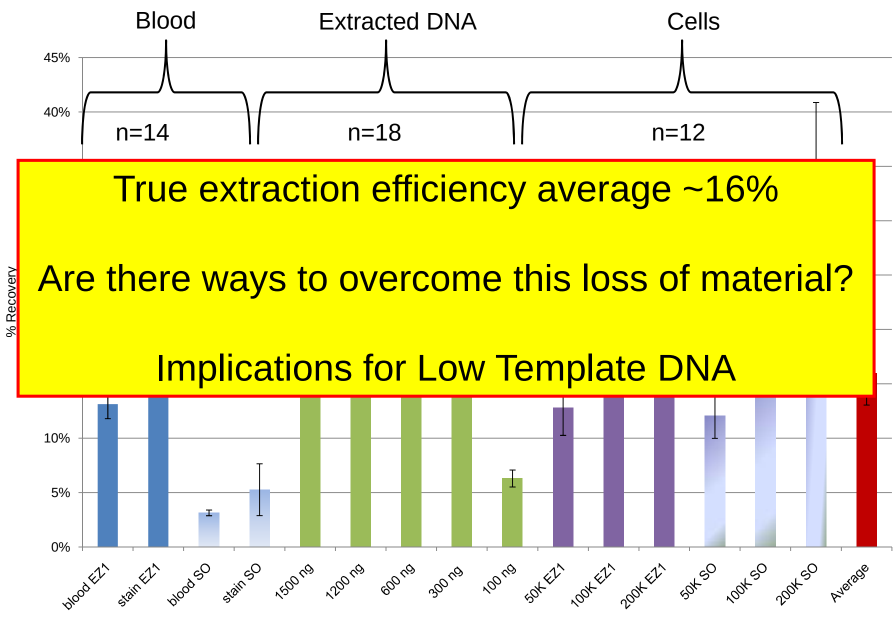
EZ1 Advanced uses magnetic separation and multiple washes to purify DNA

- Swabs & Stains: G2 Buffer and Proteinase K added to sample
- Incubated at 56°C for 15 minutes then 95°C for 5 minutes
  - Vortex periodically through incubation (~every 5 minutes)
- Blood: Total sample volume brought up to 200 µL with G2 Buffer

Modified Salt Out

- Manual extraction process
- Involves a Proteinase K digest
- Saturated Ammonium Acetate solution to separate DNA
- Absolute Ethanol wash to precipitate DNA
- Rehydrated with 100 µL TE

Extraction Efficiency Across All Samples



True extraction efficiency average ~16%

Are there ways to overcome this loss of material?

Implications for Low Template DNA

Kinship Analysis

Kristen O'Connor

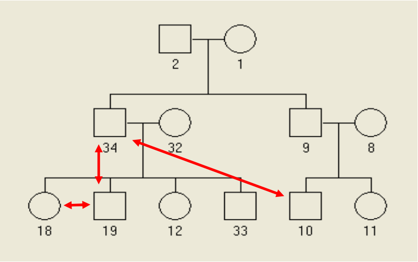
Work by Kristen Lewis O'Connor, NIST NRC Postdoc

(PhD research with Bruce Weir at University of Washington on familial search issues)

- Examine impact of additional STR loci (and other genetic markers) on addressing specific kinship questions
- Simulate likelihood ratio distributions with different sets of STR loci and different potential relationships
- Examine different software programs (and develop approaches for lab validation including investigating possible standard data sets for software testing)

Likelihood Ratios with 15 Loci

$$LR = \frac{\text{Probability of the evidence (genotypes) given a particular hypothesis}}{\text{Probability of the evidence (genotypes) given an alternative hypothesis}}$$



Sets of multigenerational families

Comparison	**LR for 34 & 19**	**LR for 18 & 19**	**LR for 34 & 10**
Parent-Child	1.28E+06		
Full Siblings		2.76E+07	
Half Siblings			6.65E-01
Cousins			
Uncle-Nephew			6.65E-01
Grandparent-Grandchild			6.65E-01

Parent/ Child

Full Sibs

Uncle/ Nephew

LR calculations were performed with GeneMarker® HIDv1.90

Benefit of Additional Loci

Likelihood Ratios with 40 Loci

	15	40	15	40	15	40
Comparison	**LR for 34 & 19**	**LR for 34 & 19**	**LR for 18 & 19**	**LR for 18 & 19**	**LR for 34 & 10**	**LR for 34 & 10**
Parent-Child	**1.28E+06**	**6.68E+16**	9.08E+05	0.00E+00	0.00E+00	0.00E+00
Full Siblings	3.22E+04	5.73E+12	**2.76E+07**	**1.57E+19**	6.07E-03	3.30E+03
Half Siblings	7.38E+03	8.63E+11	4.89E+04	4.99E+12	6.65E-01	8.98E+05
Cousins	1.95E+02	1.32E+08	8.96E+02	1.05E+09	1.52E+00	2.17E+04
Uncle-Nephew	7.38E+03	8.63E+11	4.89E+04	4.99E+12	**6.65E-01**	**8.98E+05**
Grandparent-Grandchild	7.38E+03	8.63E+11	4.89E+04	4.99E+12	6.65E-01	8.98E+05

Parent/ Child

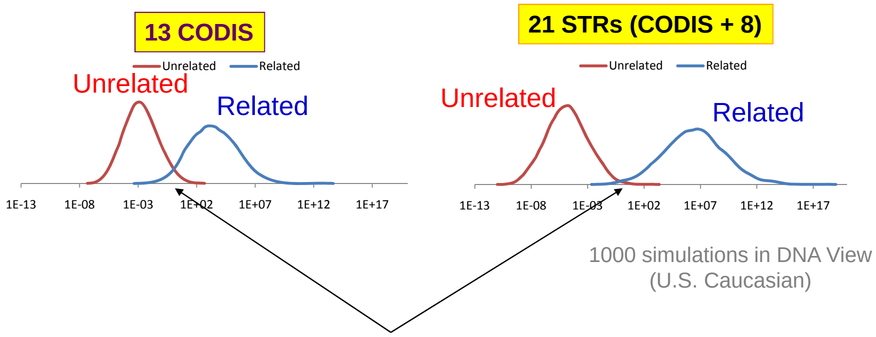
Full Sibs

Uncle/ Nephew

LR calculations were performed with GeneMarker® HIDv1.90

15 STR loci typed with commercial Identifiler kit
25 STR loci typed with an in-house NIST assay
Hill, C.R., Butler, J.M., Vallone, P.M. (2009) A 26plex autosomal STR assay to aid human identity testing. J. Forensic Sci. 54(5): 1008-1015.

LR Distributions for Full Siblings with Different Numbers of STR Loci



1000 simulations in DNA View (U.S. Caucasian)

The degree of overlap will impact potential familial search false positive results

Additional STR loci improve separation

Rapid and Direct PCR


Pete Vallone


Erica Butts

- Developing methods for faster and robust amplification of STR typing kits

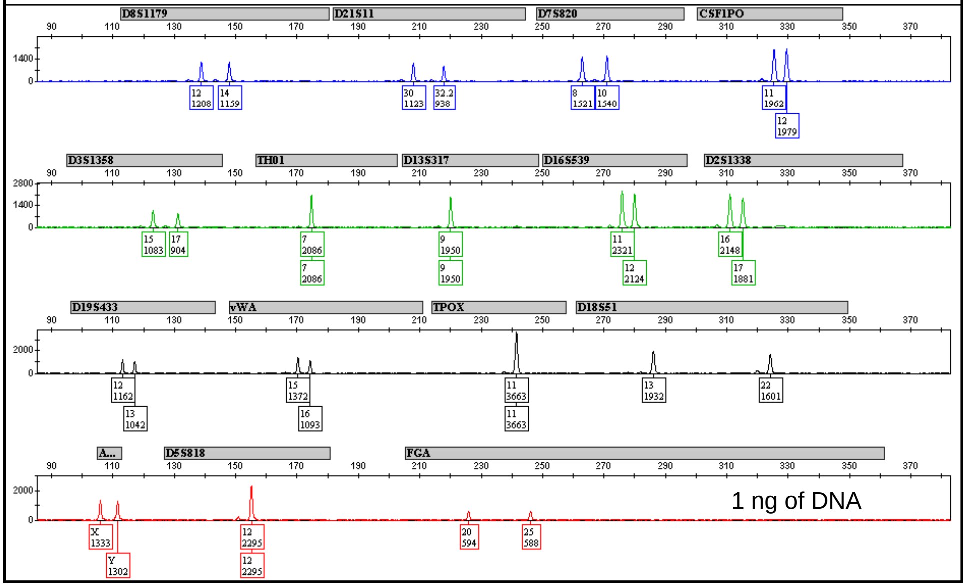
  Reduce cycling time from 3 h to 30 min
- Faster thermal cyclers
- Faster DNA polymerases

- For reference samples (non-casework)
  - 1ng of template, single source
- Apply to rapid integrated systems for DNA typing

DNA Analysis Approaches (non-integrated)

Steps Involved	*Traditional Protocols*	*Rapid Improvements (Buccal)*	*Rapid Improvements (Direct PCR)*
Collection			
Extraction	1.5 hours Manual	30 min. Automated Extraction Qiagen EZ1 Advanced	 Blood Stain
Quantitation	1.5 hours qPCR	1.5 hours	
Amplification	3.5 hours	<36 min. Rapid PCR conditions	3 hour Using PowerPlex 16 HS or Identifiler
Separation/ Detection	1 hour Capillary Electrophoresis ABI 3130xl	1 hour	1 hour
Data Interpretation	Time may vary depending on software, sample quality, and analyst expertise		
Total Time	Minimum ~7.5 hours	~4 hours	~4 hours

Full Identifiler STR Profile
Mastercycler Pro S - 19 min PCR

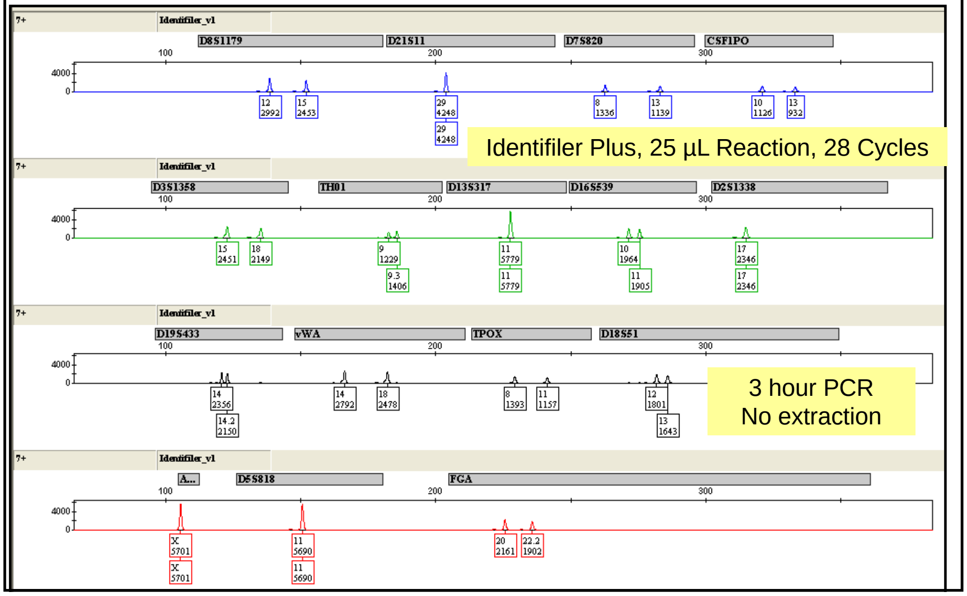


1 ng of DNA

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_FSSA.pdf

Direct PCR

1.2 mm punch from FTA paper (buccal)



Identifiler Plus, 25 µL Reaction, 28 Cycles

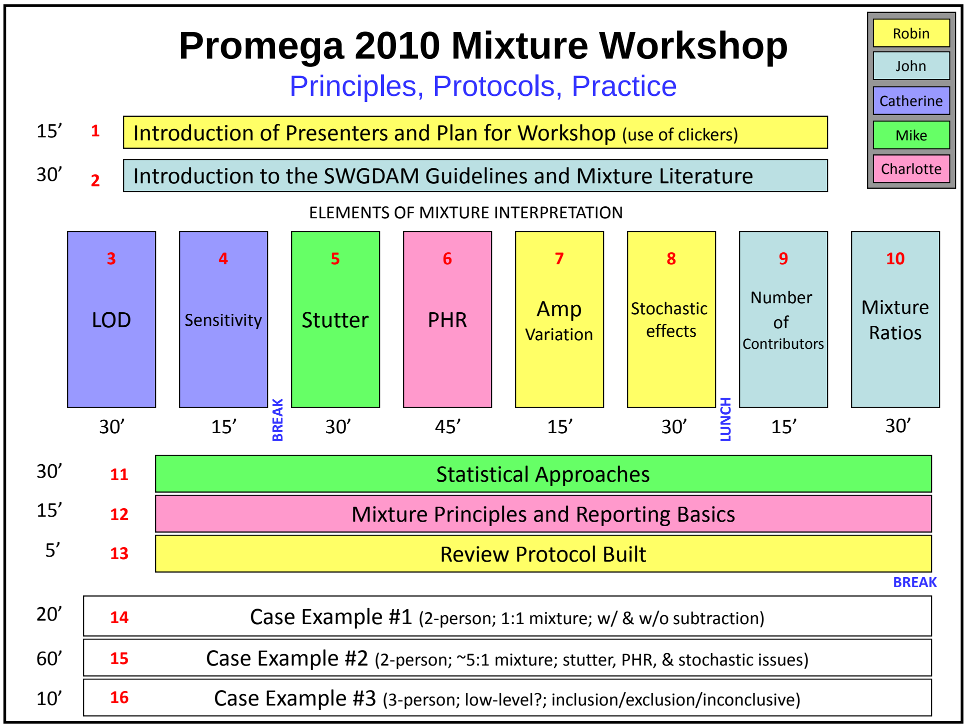
3 hour PCR
No extraction

DNA Mixture Interpretation

Mike
Coble

- Evaluate software for DNA mixture interpretation
  - True Allele
  - FSS i^3 (update)
  - USACIL (custom software)

- Developing training materials and data sets based upon SWGDAM guidelines
  - ‘Mixture of the month’
  - Mixture section on STRBase



Promega 2010 Mixture Workshop

Principles, Protocols, Practice

Robin
John
Catherine
Mike
Charlotte

15’ 1 Introduction of Presenters and Plan for Workshop (use of clickers)

30’ 2 Introduction to the SWGDAM Guidelines and Mixture Literature

ELEMENTS OF MIXTURE INTERPRETATION

3	4		5	6	7	8		9	10
LOD	Sensitivity	BREAK	Stutter	PHR	Amp Variation	Stochastic effects	LUNCH	Number of Contributors	Mixture Ratios
30’	15’		30’	45’	15’	30’		15’	30’

30’ 11 Statistical Approaches

15’ 12 Mixture Principles and Reporting Basics

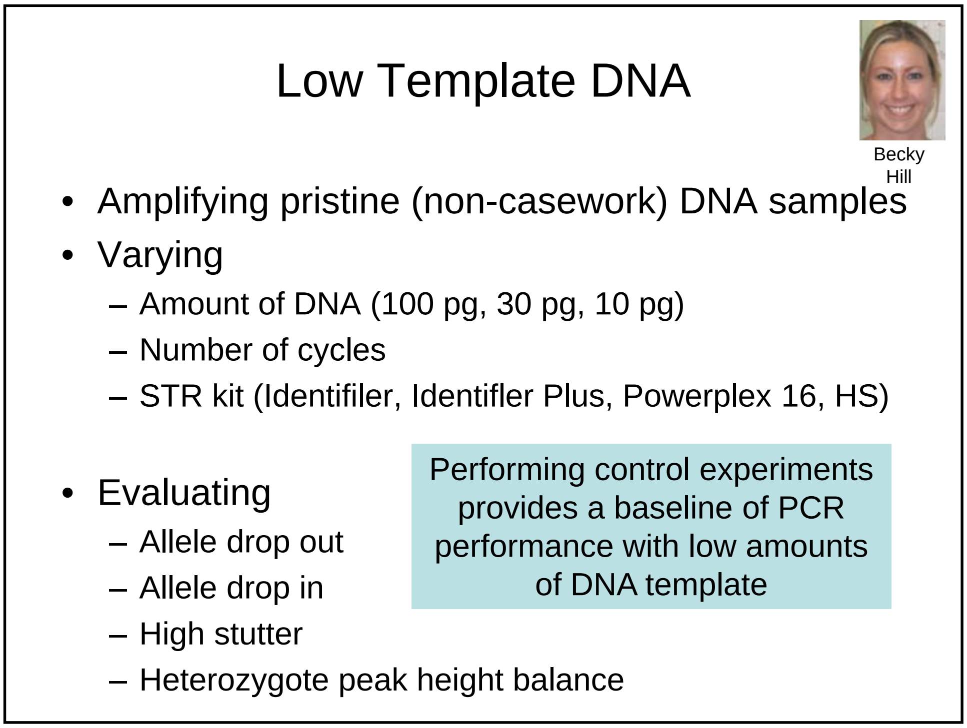
5’ 13 Review Protocol Built

BREAK

20’ 14 Case Example #1 (2-person; 1:1 mixture; w/ & w/o subtraction)

60’ 15 Case Example #2 (2-person; ~5:1 mixture; stutter, PHR, & stochastic issues)

10’ 16 Case Example #3 (3-person; low-level?; inclusion/exclusion/inconclusive)



Low Template DNA

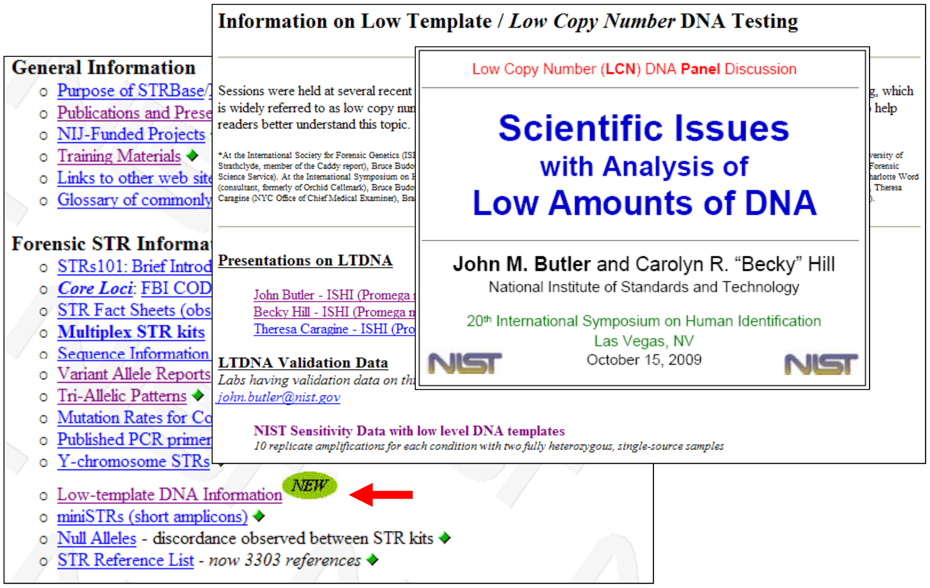
Becky Hill

- Amplifying pristine (non-casework) DNA samples
- Varying
  - Amount of DNA (100 pg, 30 pg, 10 pg)
  - Number of cycles
  - STR kit (Identifiler, Identifler Plus, Powerplex 16, HS)
- Evaluating
  - Allele drop out
  - Allele drop in
  - High stutter
  - Heterozygote peak height balance

Performing control experiments provides a baseline of PCR performance with low amounts of DNA template

New STRBase Website on LT-DNA (LCN)

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



General Information

- Purpose of STRBase/
- Publications and Prese
- NIJ-Funded Projects
- Training Materials ◆
- Links to other web site
- Glossary of commonly

Forensic STR Informa

- STRs101: Brief Introd
- *Core Loci*: FBI COD
- STR Fact Sheets (obs
- **Multiplex STR kits**
- Sequence Information
- Variant Allele Reports
- Tri-Allelic Patterns ◆
- Mutation Rates for Co
- Published PCR primer
- Y-chromosome STRs
- Low-template DNA Information NEW
- miniSTRs (short amplicons) ◆
- Null Alleles - discordance observed between STR kits ◆
- STR Reference List - *now 3303 references* ◆

Information on Low Template / *Low Copy Number* DNA Testing

Sessions were held at several recent … is widely referred to as low copy num… readers better understand this topic. …g, which …help

Presentations on LTDNA

John Butler - ISHI (Promega
Becky Hill - ISHI (Promega m
Theresa Caragine - ISHI (Pro

LTDNA Validation Data

Labs having validation data on th
john.butler@nist.gov

NIST Sensitivity Data with low level DNA templates

10 replicate amplifications for each condition with two fully heterozygous, single-source samples

Low Copy Number (**LCN**) DNA **Panel** Discussion

Scientific Issues with Analysis of Low Amounts of DNA

John M. Butler and Carolyn R. "Becky" Hill
National Institute of Standards and Technology

20th International Symposium on Human Identification
Las Vegas, NV
October 15, 2009

Complete Set of NIST Sensitivity Data Available on New LT-DNA Website

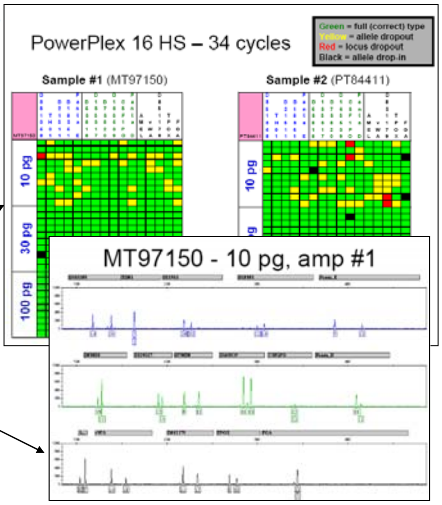
http://www.cstl.nist.gov/biotech/strbase/LTDNA.htm

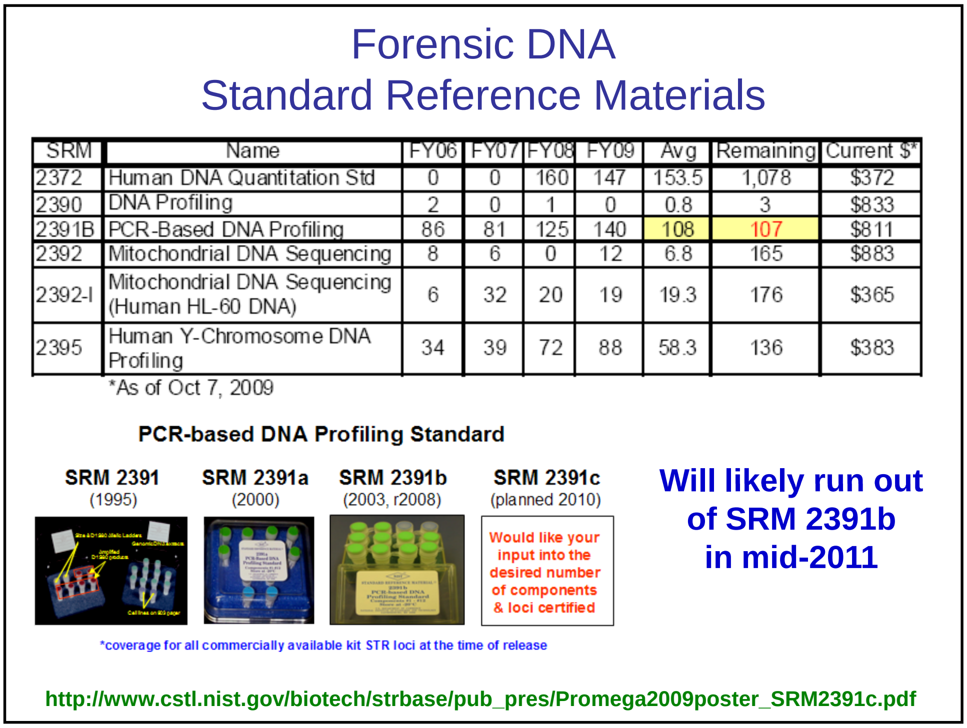
NIST Sensitivity Data with low level DNA templates

10 replicate amplifications for each condition with two fully heterozygous, single-source samples

Click on links to see summaries and DNA profiles observed

STR kit - PCR conditions	Sample 1	Sample 2
Identifiler - 28 cycles	100 pg	100 pg
	30 pg	30 pg
	10 pg	10 pg
Identifiler - 31 cycles	**100 pg**	**100 pg**
	30 pg	**30 pg**
	10 pg	**10 pg**
PowerPlex 16 HS - 31 cycles	100 pg	100 pg
	30 pg	30 pg
	10 pg	10 pg
PowerPlex 16 HS - 34 cycles	**100 pg**	**100 pg**
	30 pg	**30 pg**
	10 pg	**10 pg**





Forensic DNA Standard Reference Materials

SRM	Name	FY06	FY07	FY08	FY09	Avg	Remaining	Current $*
2372	Human DNA Quantitation Std	0	0	160	147	153.5	1,078	$372
2390	DNA Profiling	2	0	1	0	0.8	3	$833
2391B	PCR-Based DNA Profiling	86	81	125	140	108	107	$811
2392	Mitochondrial DNA Sequencing	8	6	0	12	6.8	165	$883
2392-I	Mitochondrial DNA Sequencing (Human HL-60 DNA)	6	32	20	19	19.3	176	$365
2395	Human Y-Chromosome DNA Profiling	34	39	72	88	58.3	136	$383

*As of Oct 7, 2009

PCR-based DNA Profiling Standard

SRM 2391 (1995)

SRM 2391a (2000)

SRM 2391b (2003, r2008)

SRM 2391c (planned 2010)

Would like your input into the desired number of components & loci certified

Will likely run out of SRM 2391b in mid-2011

*coverage for all commercially available kit STR loci at the time of release

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2009poster_SRM2391c.pdf

SRM 2391c Plans

- 6 components (different DNA samples)
- Inclusion of a mixture sample
- Genomic DNA + DNA spotted on paper
- Will certify both autosomal STRs and Y-STRs
- Will replace both SRM 2391b and SRM 2395
  - Goal to have lower cost and single STR SRM

- Have typed >25 DNA candidate samples to assess allele ranges and combinations
- Have studied different formats
  - buccal swab DNA, etc.

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_FSSA.pdf

Workshops and Training

So far in 2010…

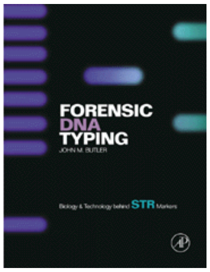
- 10 articles published
- Over 20 presentations given

- DNA as a Biometric Workshop (AAFS 2010)
  - Pete Vallone
- Mixture workshops (at Promega 2010 and AAFS 2011)
  - Mike Coble and John Butler

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

The Expansion of *Forensic DNA Typing*

1st Edition



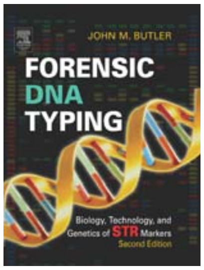
Jan 2001

335 pp.

17 chapters

John Butler

2nd Edition



Feb 2005

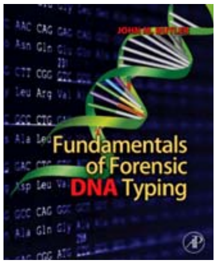
688 pp.

24 chapters

Chinese Translation
(2007) Y. Hou, translator

Japanese Translation
(2009) Y. Fukuma, translator

3rd Edition

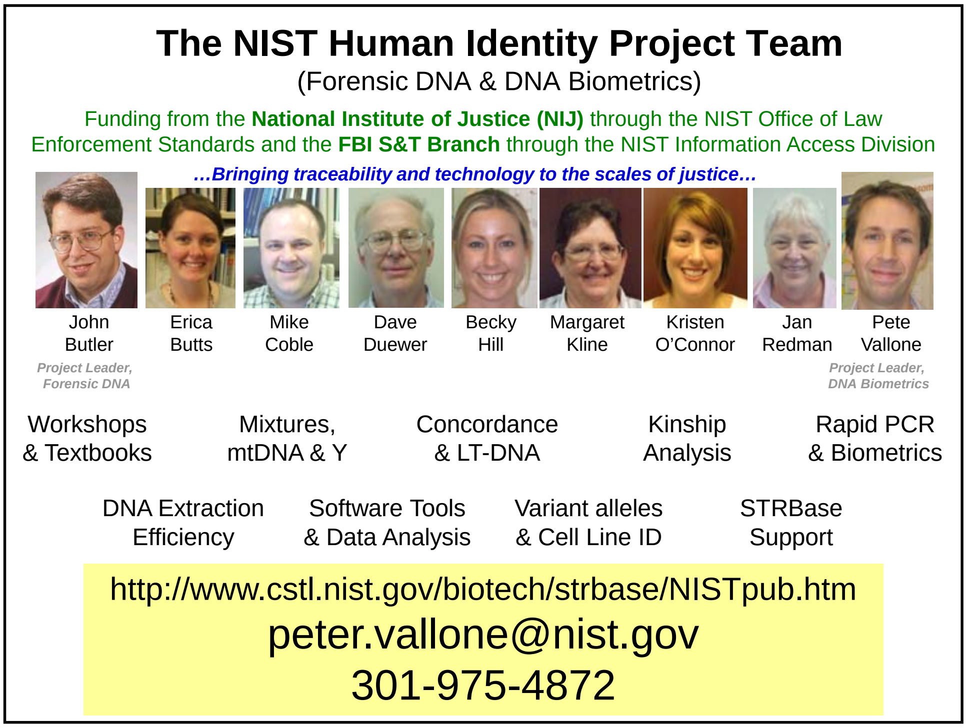


Sept 2009

Fundamentals
18 chapters (504 pp.)

Advanced Topics
25 chapters (~600 pp.)

Planned for 2011



The NIST Human Identity Project Team

(Forensic DNA & DNA Biometrics)

Funding from the **National Institute of Justice (NIJ)** through the NIST Office of Law Enforcement Standards and the **FBI S&T Branch** through the NIST Information Access Division

…Bringing traceability and technology to the scales of justice…

John Butler
Project Leader, Forensic DNA

Erica Butts

Mike Coble

Dave Duewer

Becky Hill

Margaret Kline

Kristen O'Connor

Jan Redman

Pete Vallone
Project Leader, DNA Biometrics

Workshops & Textbooks

Mixtures, mtDNA & Y

Concordance & LT-DNA

Kinship Analysis

Rapid PCR & Biometrics

DNA Extraction Efficiency

Software Tools & Data Analysis

Variant alleles & Cell Line ID

STRBase Support

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

peter.vallone@nist.gov

301-975-4872