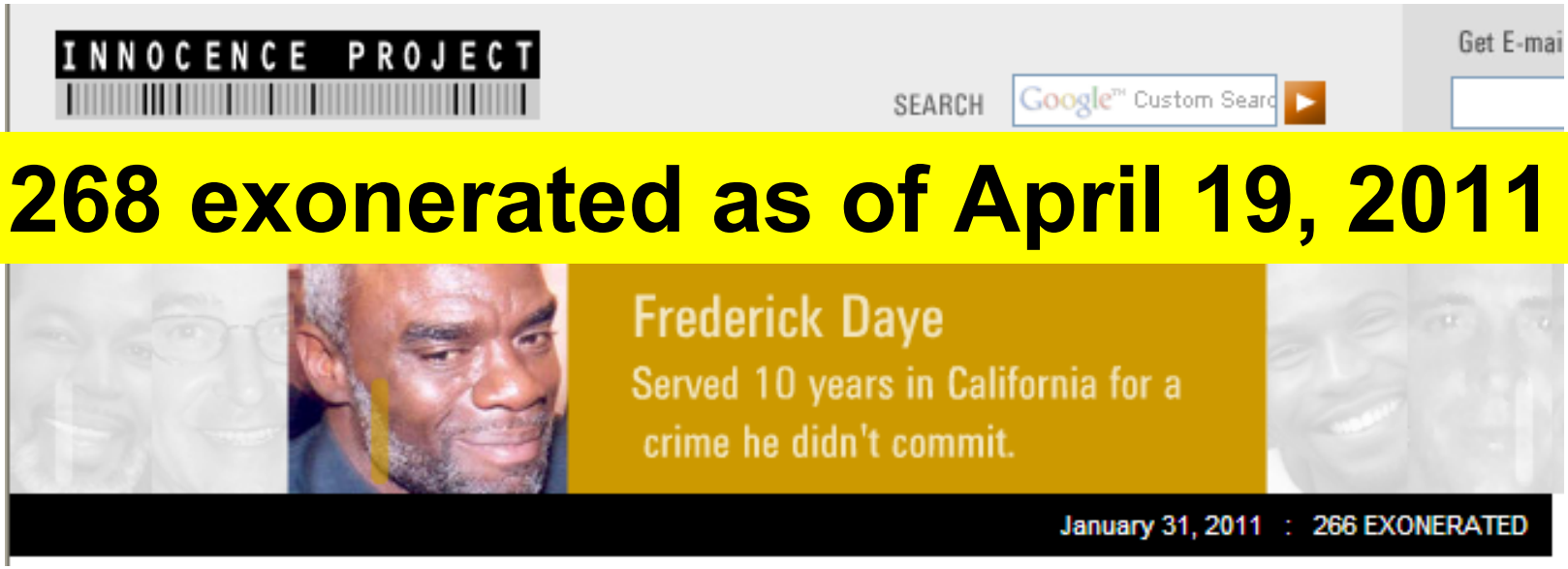
DNA Databases: Uses and Issues

John M. Butler

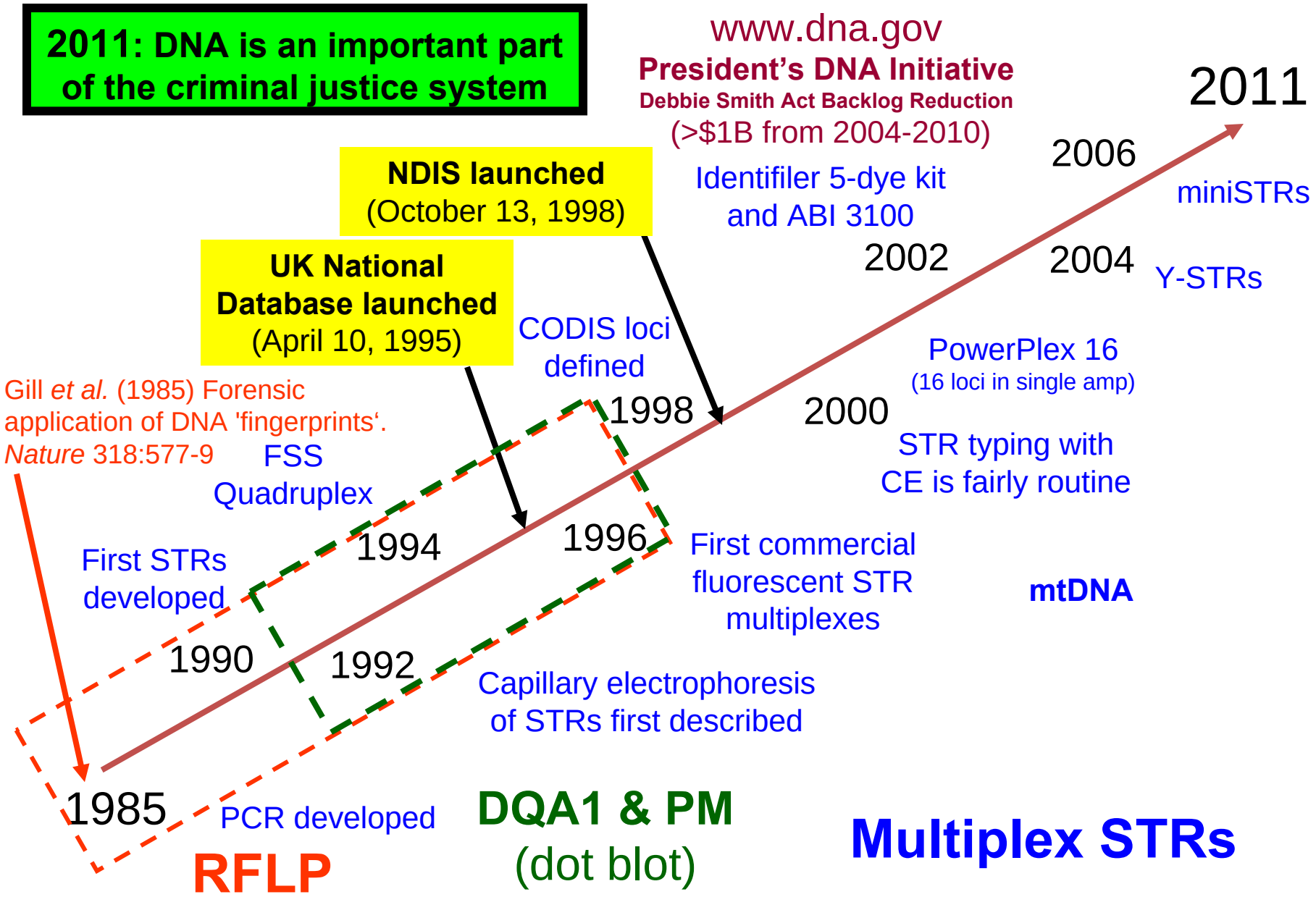
National Institute of Standards and Technology (NIST)

Quality Results are Essential in Forensic DNA Testing

- **DNA results impact lives** – the guilty can be implicated in a crime and the innocent can be exonerated
- Scientific attacks against the science behind DNA testing are rare in court now. Rather **the focus is on demonstrating that quality results were obtained**.
- **DNA databases involve comparisons** of DNA profiles analyzed at different times or in different locations

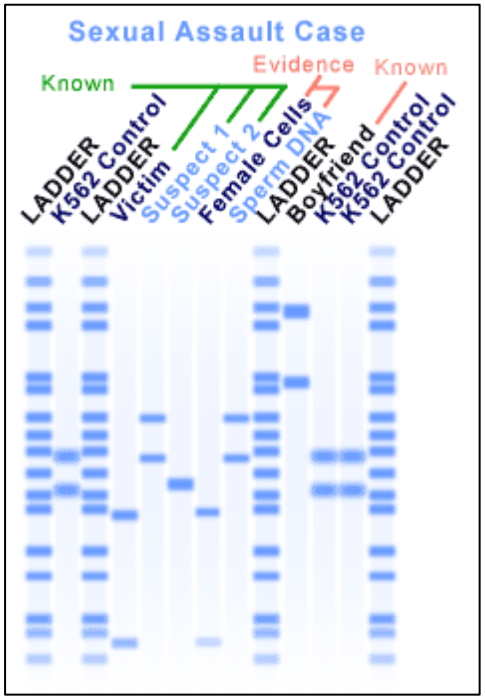


Historical Perspective on DNA Typing



The DNA Field Moves Forward…

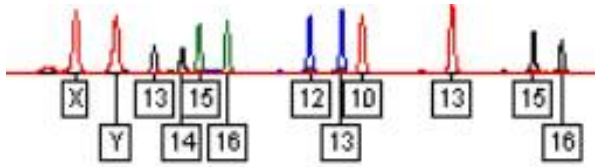
The Past



http://www.bioteach.ubc.ca/MolecularBiology/DNAfingerprint/

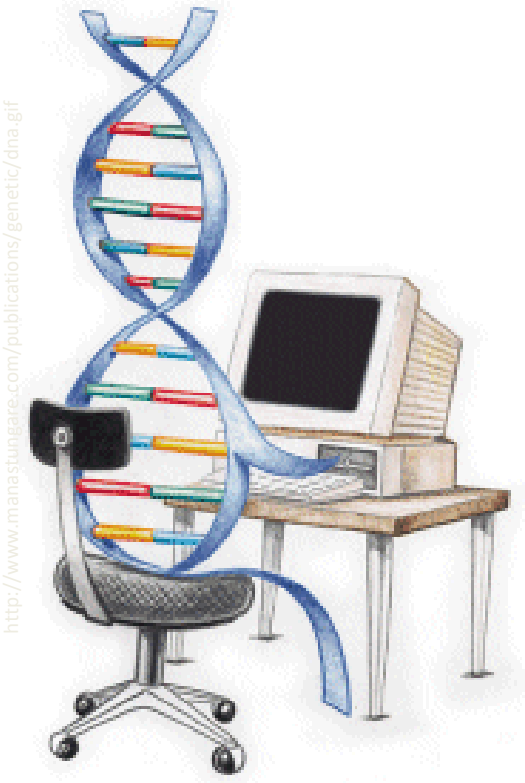
RFLP

The Present



STRs

The Future



http://www.manastungare.com/publications/genetic/dna.gif

DNA Testing Requires a Reference Sample

A DNA profile by itself is fairly useless because it has no context…

DNA analysis for identity only works by comparison – **you need a reference sample**

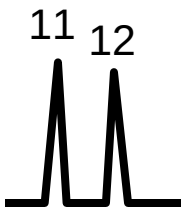
Crime Scene Evidence compared to **Suspect(s)** (Forensic Case)
Child compared to **Alleged Father** (Paternity Case)
Victim's Remains compared to **Biological Relative** (Mass Disaster ID)
Soldier's Remains compared to **Direct Reference Sample** (Armed Forces ID)

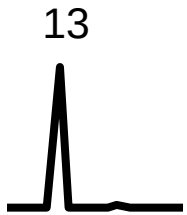
The Three Possible Outcomes
of Evidence Examination

"Suspect"
Known (K) Sample

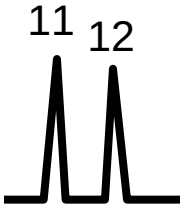
"Evidence"
Question (Q) Sample

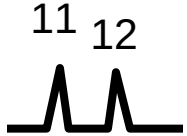
- **Exclusion** (no match)



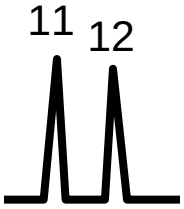


- **Non-exclusion**
  - "Match" or "inclusion"

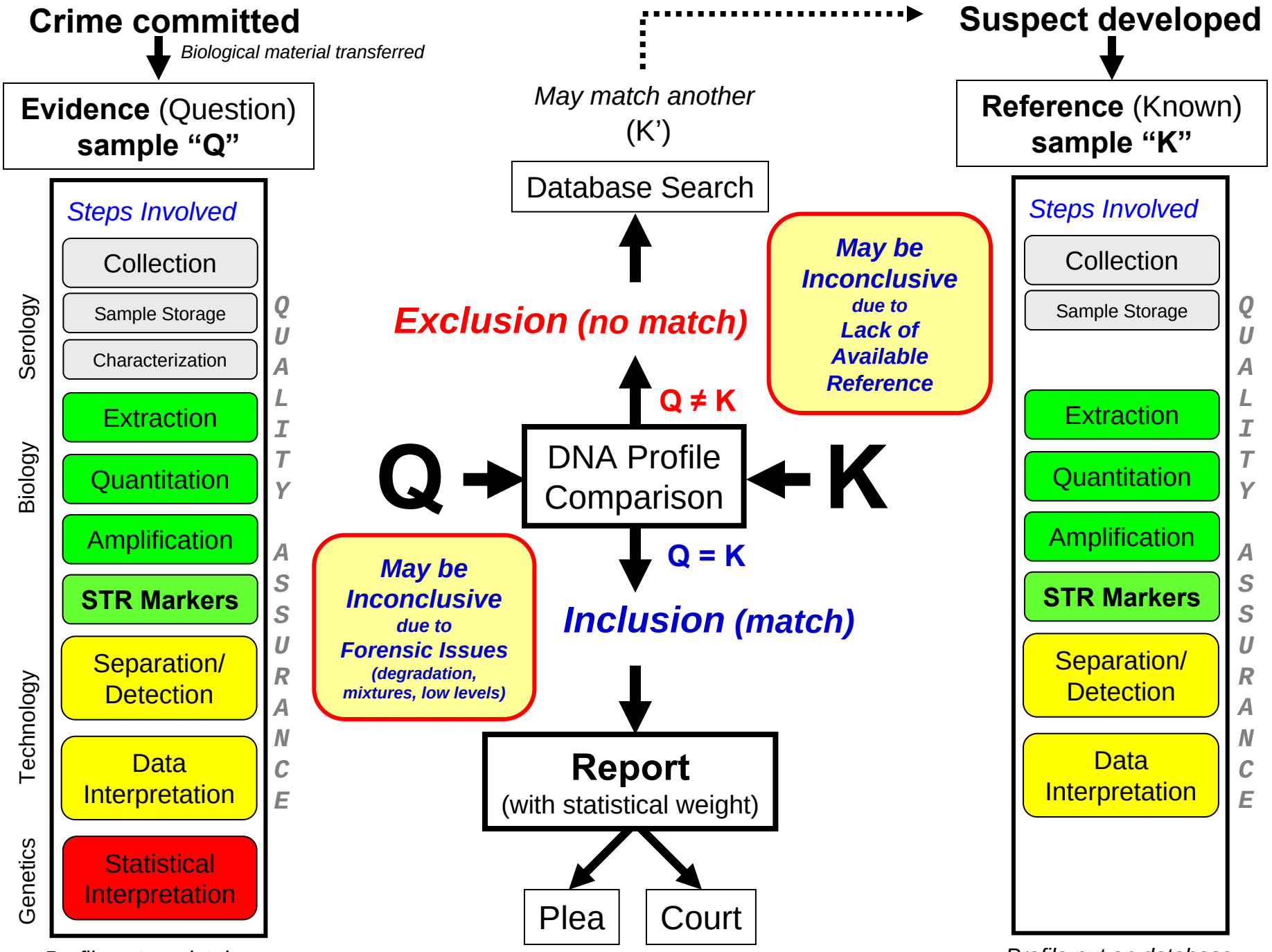




- **Inconclusive result**



No result
(or a complex mixture)



John M. Butler (2009) *Fundamentals of Forensic DNA Typing*, Figure 1.3

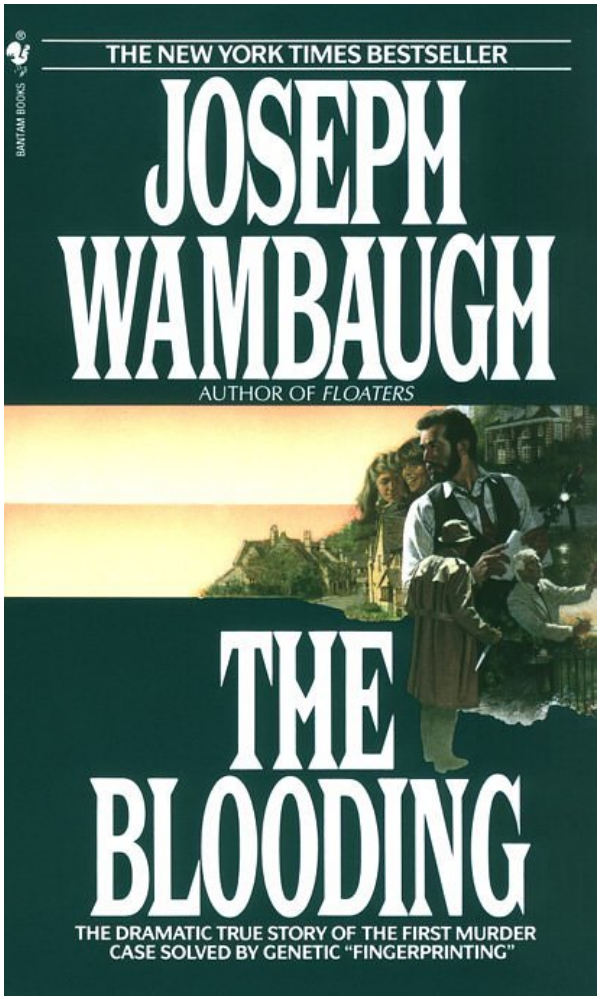
Applications for DNA Testing

- Crime solving – **matching suspect with evidence…**
- Accident victims – **after airplane crashes…**
- Soldiers in war – **who is the "unknown" soldier…**
- Paternity testing – **who is the father…**
- Immigration testing **– are two people related…**
- Missing persons investigations – **whose remains…**
- Convicted felons databases – **cases solved…**

Involves generation of DNA profiles usually with the same core STR (short tandem repeat) markers and then MATCHING TO REFERENCE SAMPLE

Lessons from the First Case Involving DNA Testing

John M. Butler (2009) *Fundamentals of Forensic DNA Typing*, D.N.A. Box 1.2



Describes the first use of DNA (in 1986) to solve a double rape-homicide case in England; about 5,000 men asked to give blood or saliva to compare to crime stains

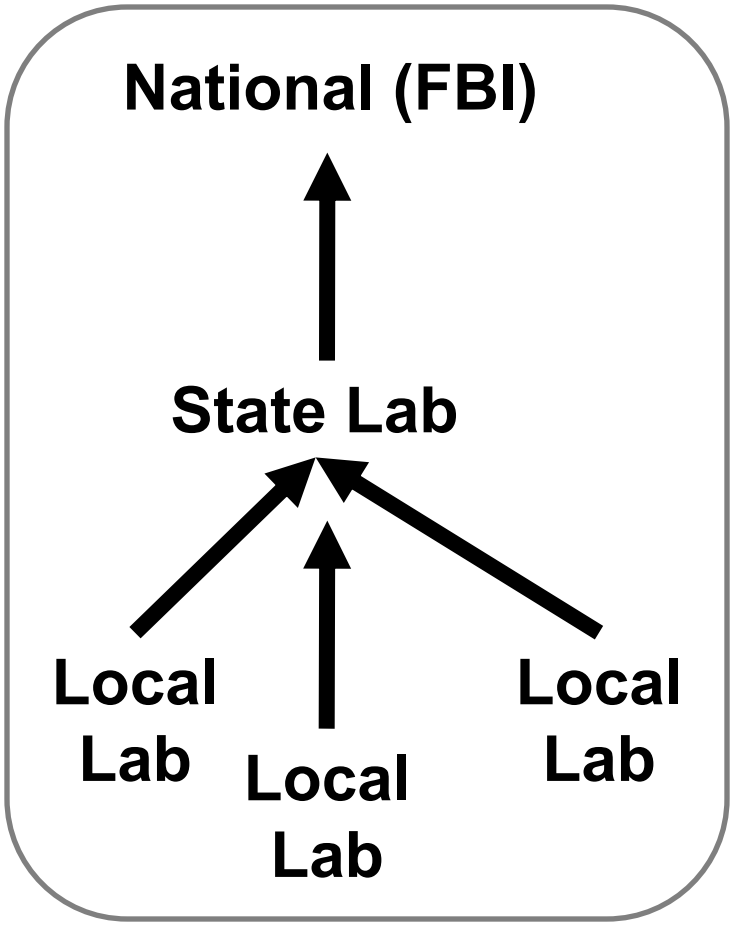
- Connection of two crimes (1983 and 1986)
- Use of DNA database to screen for perpetrator (DNA only done on 10% with same blood type as perpetrator)
- Exoneration of an innocent suspect
- DNA was an investigative tool – did not solve the case by itself (confession of accomplice)

A local baker, Colin Pitchfork, was arrested and his DNA profile matched with the semen from both murders. In 1988 he was sentenced to life for the two murders.

U.S. National DNA Database

National DNA Index System (NDIS)

CODIS Levels



- CODIS = Combined DNA Index System
- **190 public labs (government)**
  - 136 local
  - 54 state
- About **12 private labs** contribute data that must be reviewed and approved by public labs prior to upload

Growth of DNA Databases

- Expanded laws now enable more offenders to be included (25 states collect from arrestees)
  - Has contributed to sample backlogs
- Have benefited from significant federal funding since 2004 (>$1 billion for backlog reduction)
- Have effectively locked technology with core STR markers used to generate DNA profiles that now number in the millions

California State DNA Sample Backlog

CA adds about 20,000 samples per month

Month	**November 2006**	**July 2009**	**November 2010**
Starting Backlog	221,052	61,611	**39,651**
Ending Backlog	197,227	60,815	**41,679**
Total Offender Profiles in SDIS	662,542	1,294,314	1,660,025
Total Forensic Unknowns in SDIS	14,813	26,887	35,800
Hits (that month)	201	317	343
Total Hits (cumulative)	3346	9701	14,925

For most recent data, see http://ag.ca.gov/bfs/pdf/Monthly.pdf

Advocates for DNA Funding and Expansion

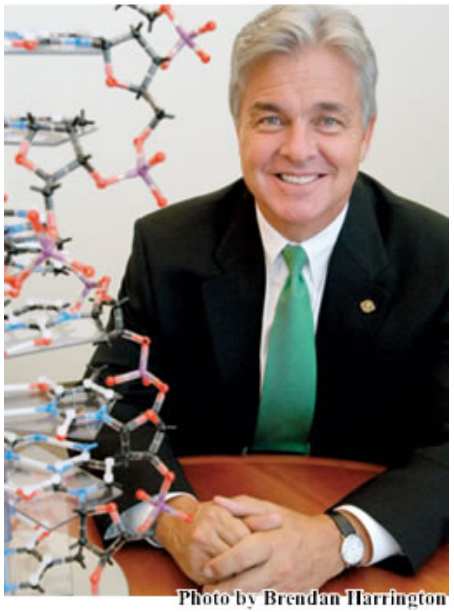
Debbie Smith

http://www.ojp.usdoj.gov/ovc/ncvrw/2005/images/d_smith_photo.jpg

Victim

Mitch Morrissey

http://www.denverda.org/images/DNApic.jpg



Prosecutor

Kirk Bloodsworth

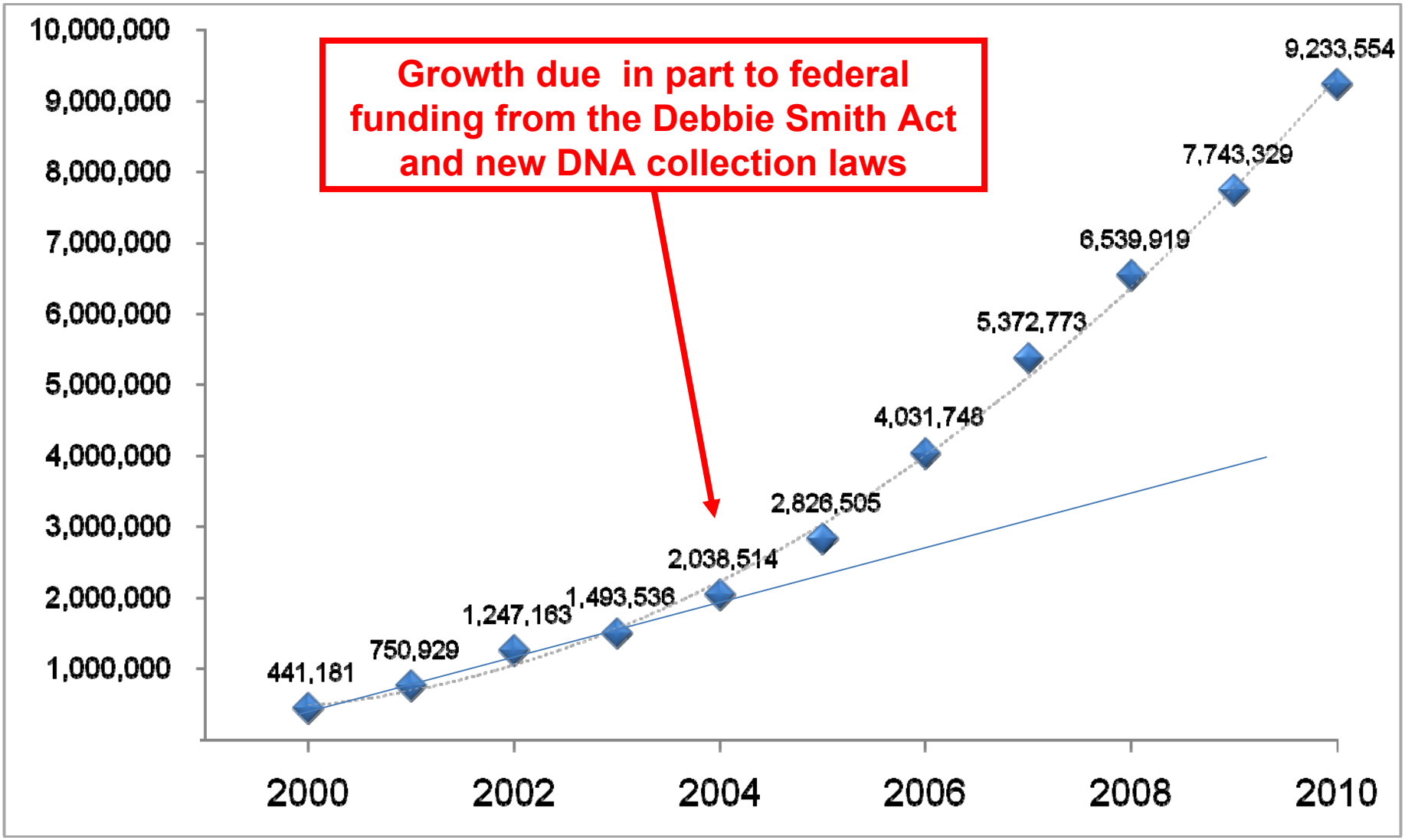
http://www.waynesburg.edu/files/Bloodsworth%20Headshot.jpg

Exoneree
(Innocence Project)

Debbie Smith Act of 2004 and Reauthorization Act of 2008 has provided **$150M per year (2004-2014) for federal funds to state and local labs** for backlog reduction

Number of Offender DNA Profiles

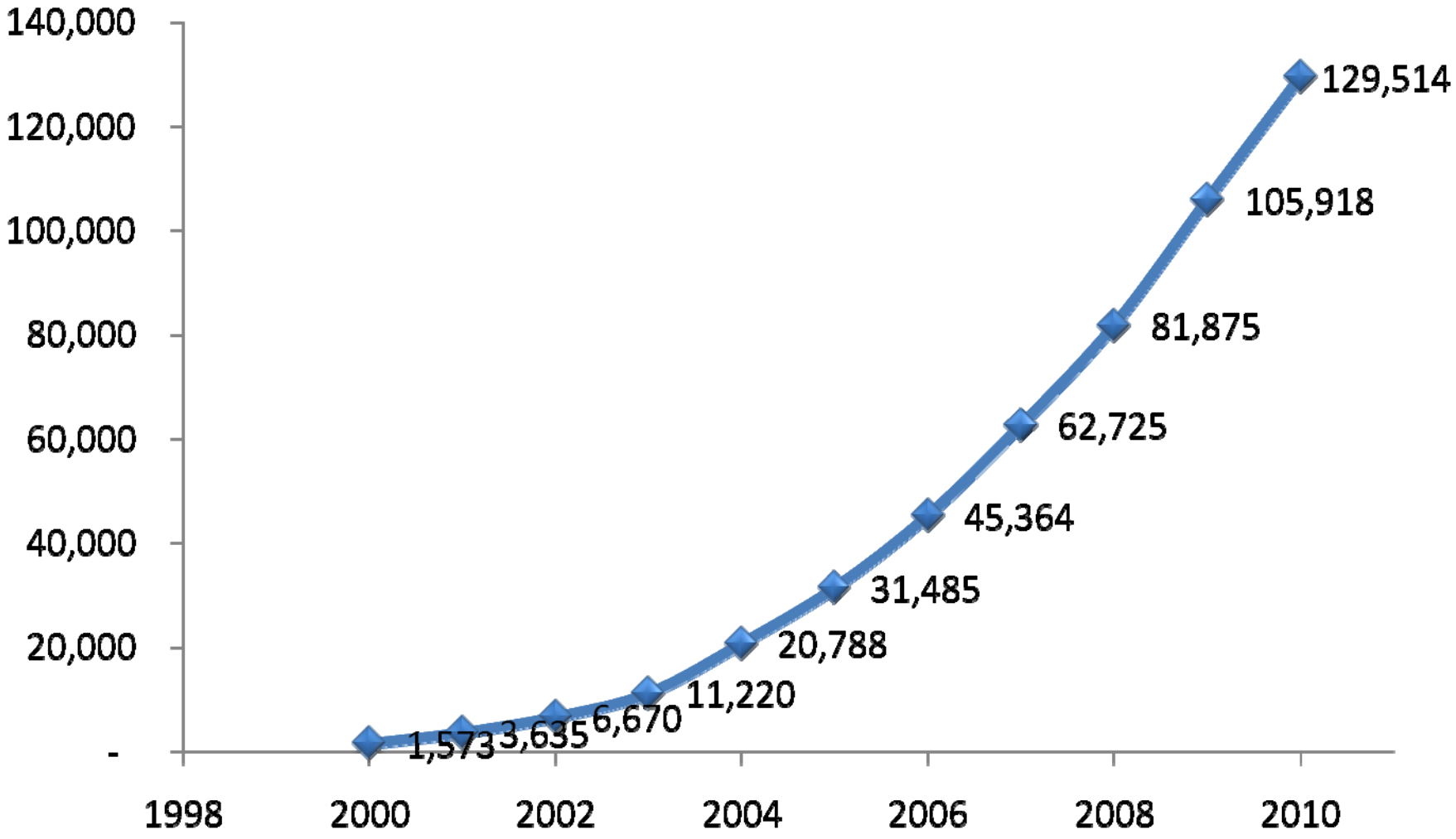
in the U.S. National DNA Database



Source: FBI Laboratory's CODIS Unit

Number of Investigations Aided

in the U.S. National DNA Database



Source: FBI Laboratory's CODIS Unit

Steps in Forensic DNA Analysis

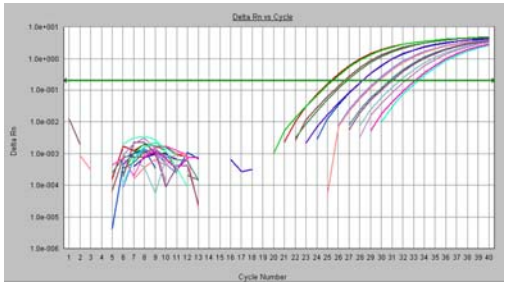
Usually 1-2 day process (a minimum of ~8 hours)

Blood Stain Buccal swab

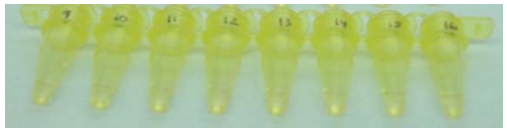
Sample Collection & Storage

Biology

DNA Extraction

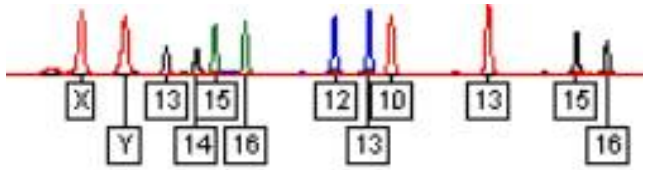


DNA Quantitation



Multiplex PCR Amplification

DNA separation and sizing



Technology

STR Typing

Interpretation of Results

Genetics

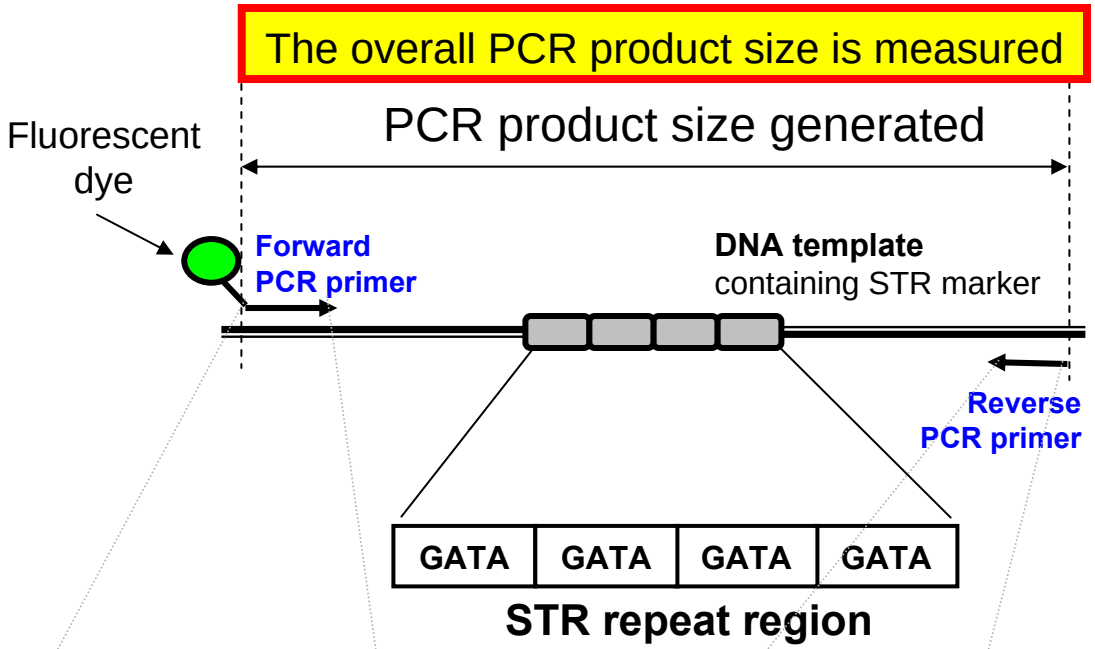
Statistics Calculated

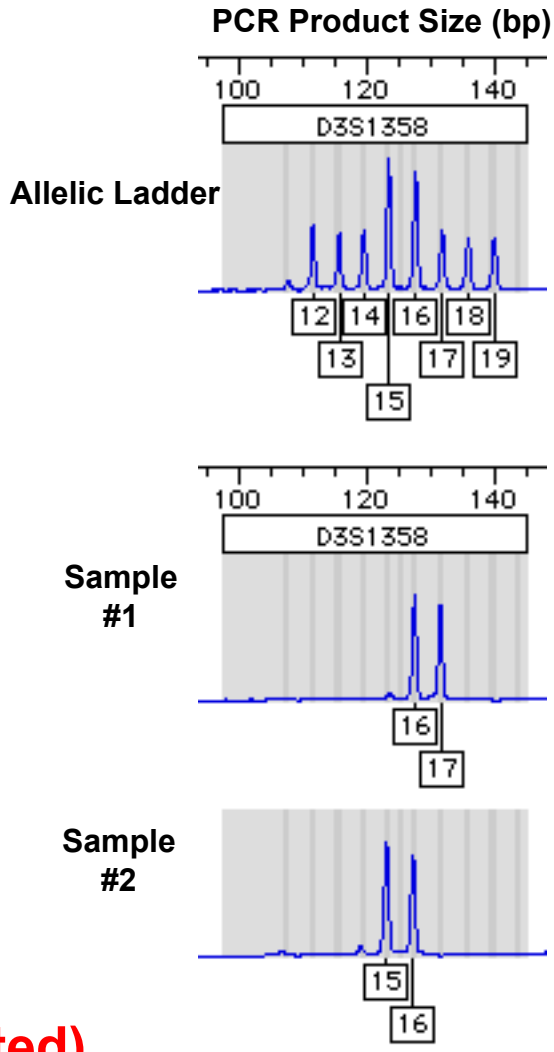
DNA Database search
Paternity test
Reference sample

Applied Use of Information

Short Tandem Repeat (STR) Markers

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

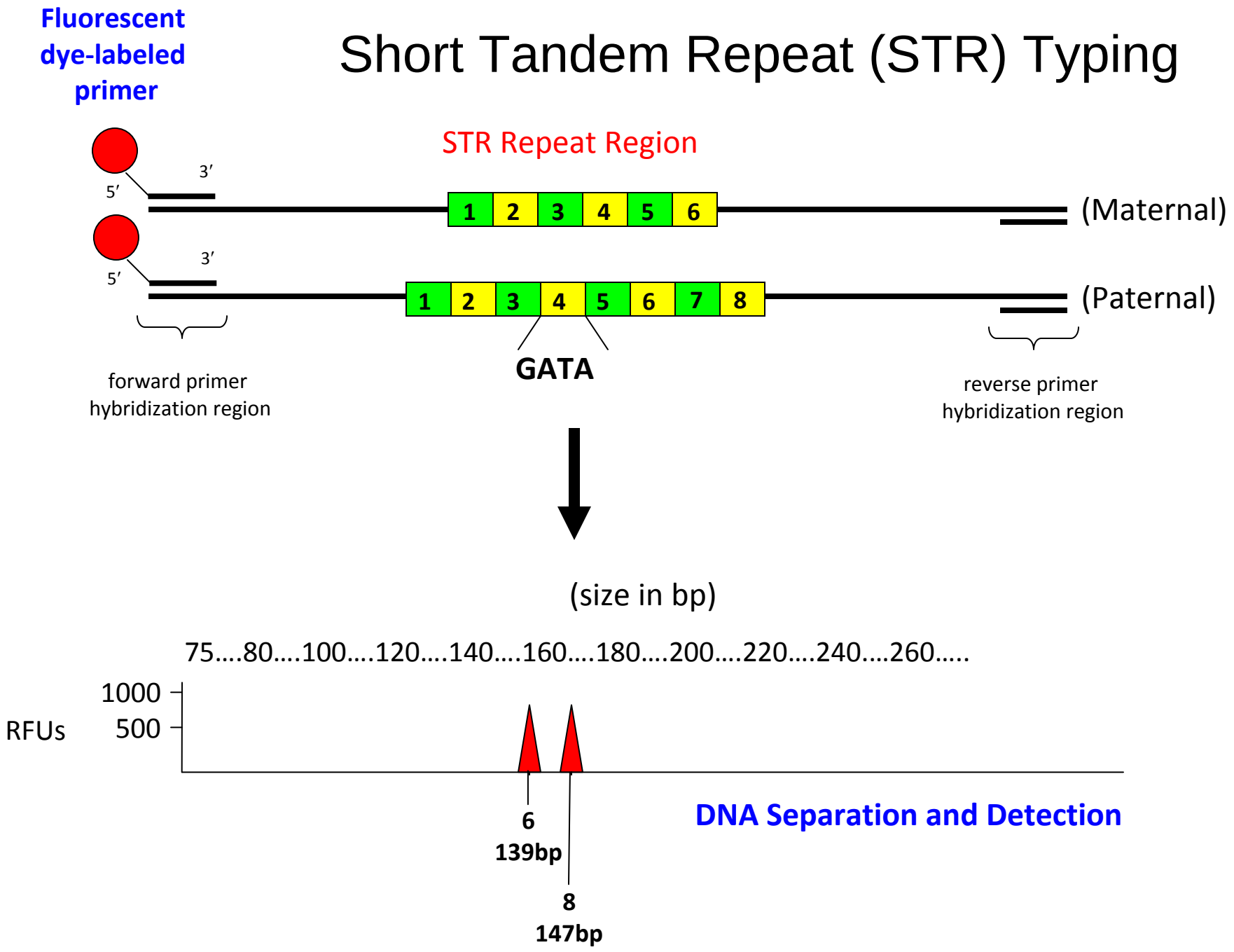




TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGA
AGACAGGTG**GATAGATAGATAGATAGATAGATAGATA**
GATAGATAGATAGATATCATTGAAAGACAAAACAGAGA
TGGATGATAGATACATGCTTACAGATGCACAC

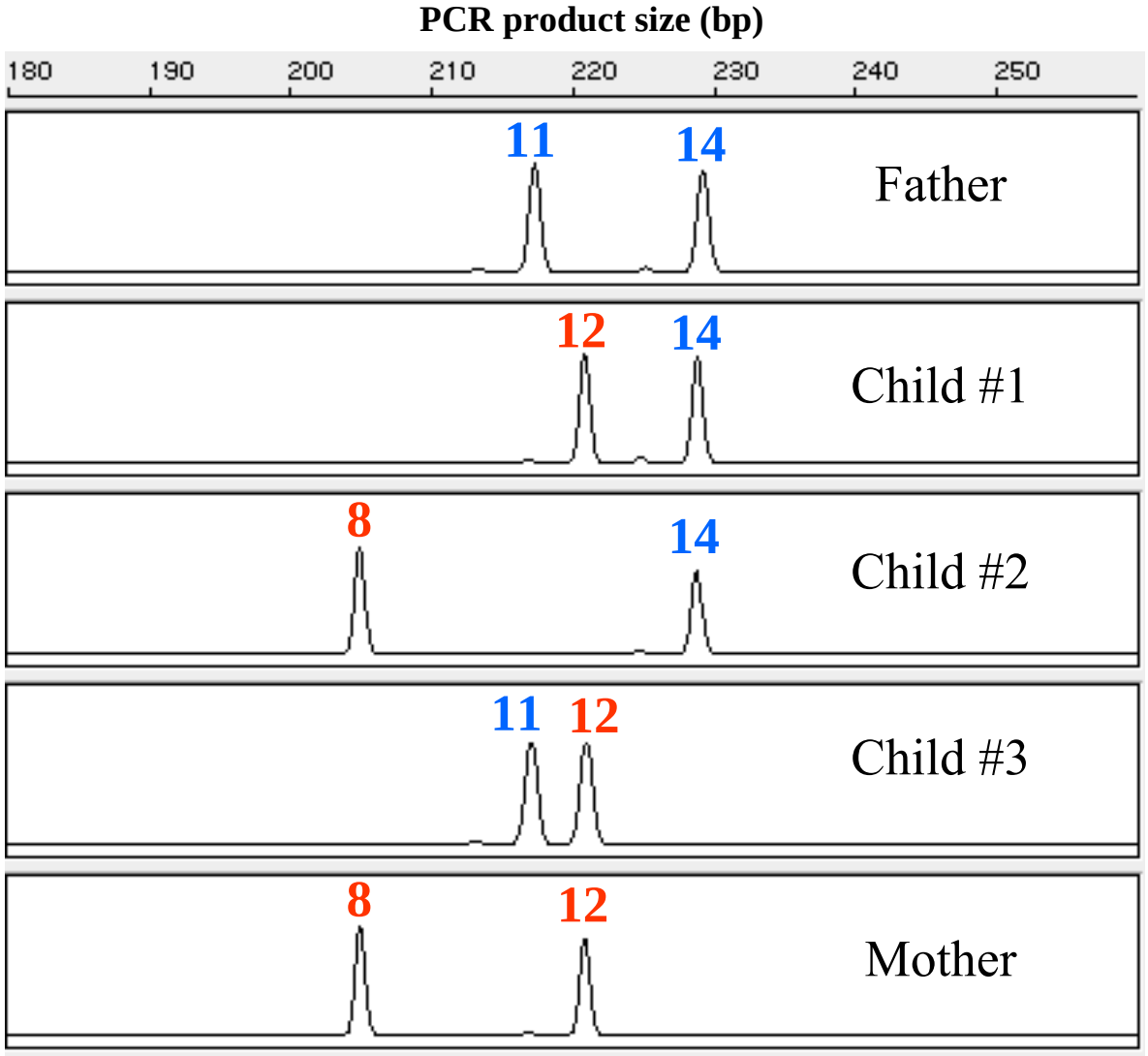
= 11 GATA repeats ("11" is all that is reported)

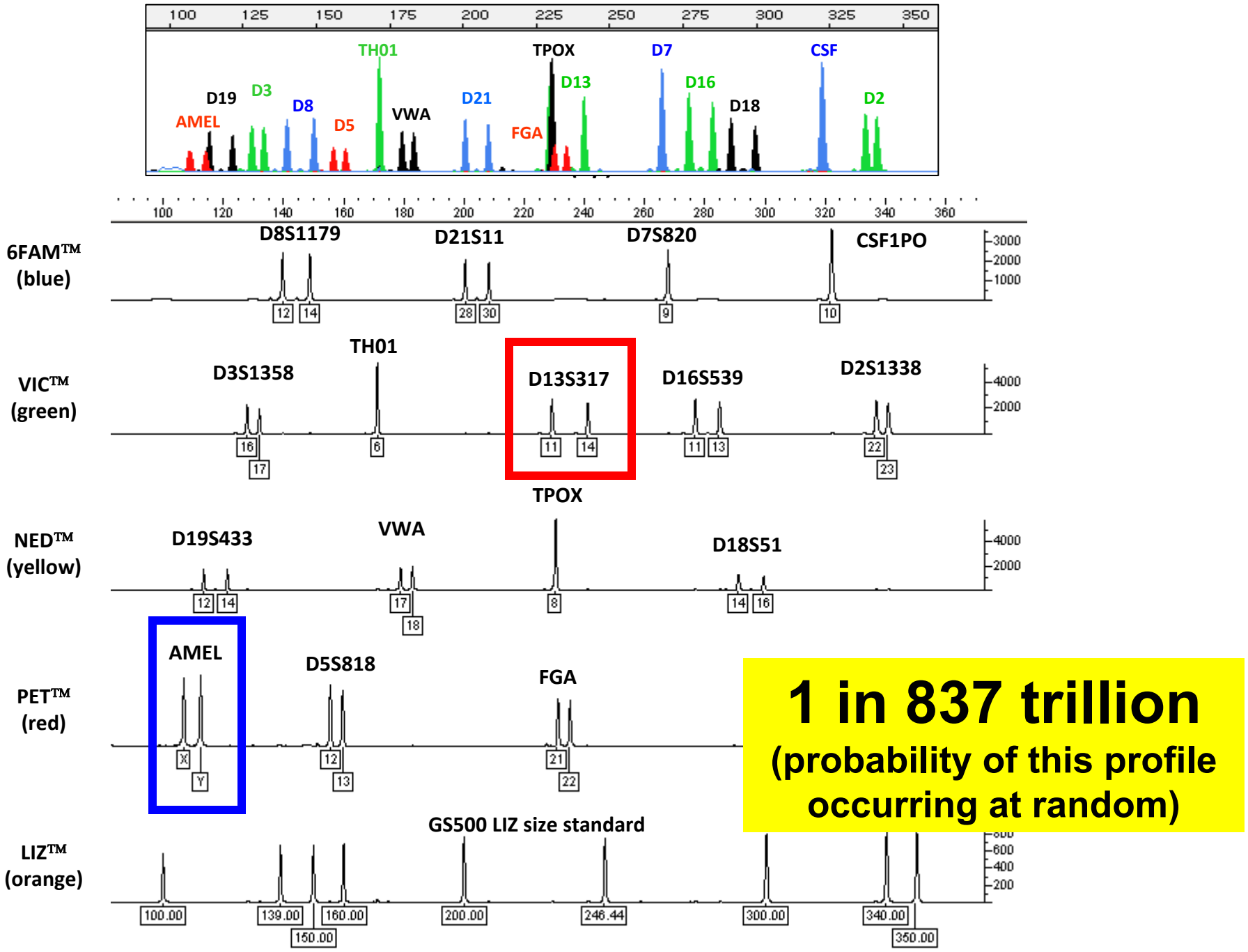
Short Tandem Repeat (STR) Typing



PATERNITY TESTING

Family Inheritance of STR Alleles (D13S317)

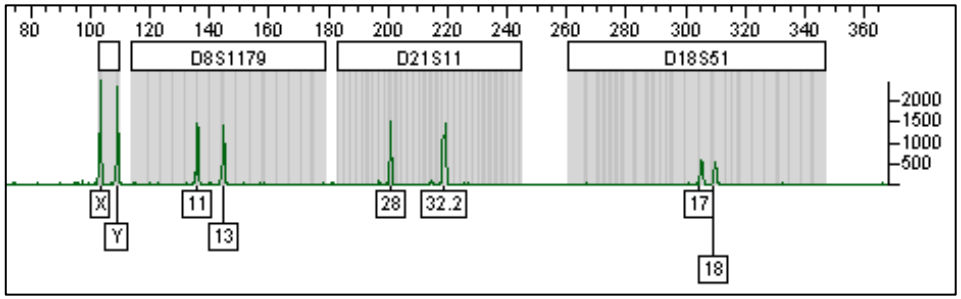



100 125 150 175 200 225 250 275 300 325 350
AMEL
D19
D3
D8
D5
TH01
VWA
D21
FGA
TPOX
D13
D7
D16
D18
CSF
D2
100 120 140 160 180 200 220 240 260 280 300 320 340 360
6FAM™ (blue)
D8S1179
D21S11
D7S820
CSF1PO
12 14
28 30
9
10
3000
2000
1000
VIC™ (green)
D3S1358
TH01
D13S317
D16S539
D2S1338
16 17
6
11 14
11 13
22 23
4000
2000
NED™ (yellow)
TPOX
D19S433
VWA
D18S51
12 14
17 18
8
14 16
4000
2000
PET™ (red)
AMEL
D5S818
FGA
X Y
12 13
21 22
LIZ™ (orange)
GS500 LIZ size standard
100.00
139.00
150.00
160.00
200.00
246.44
300.00
340.00
350.00
600
400
200
1 in 837 trillion
(probability of this profile occurring at random)

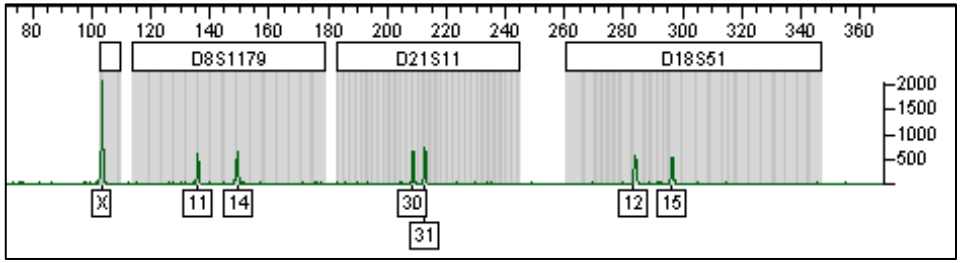
STR Results

- Individuals will differ from one another in terms of their STR profile
- STR genotype can then be put into an alpha numeric form for search on a DNA database

Individual #1



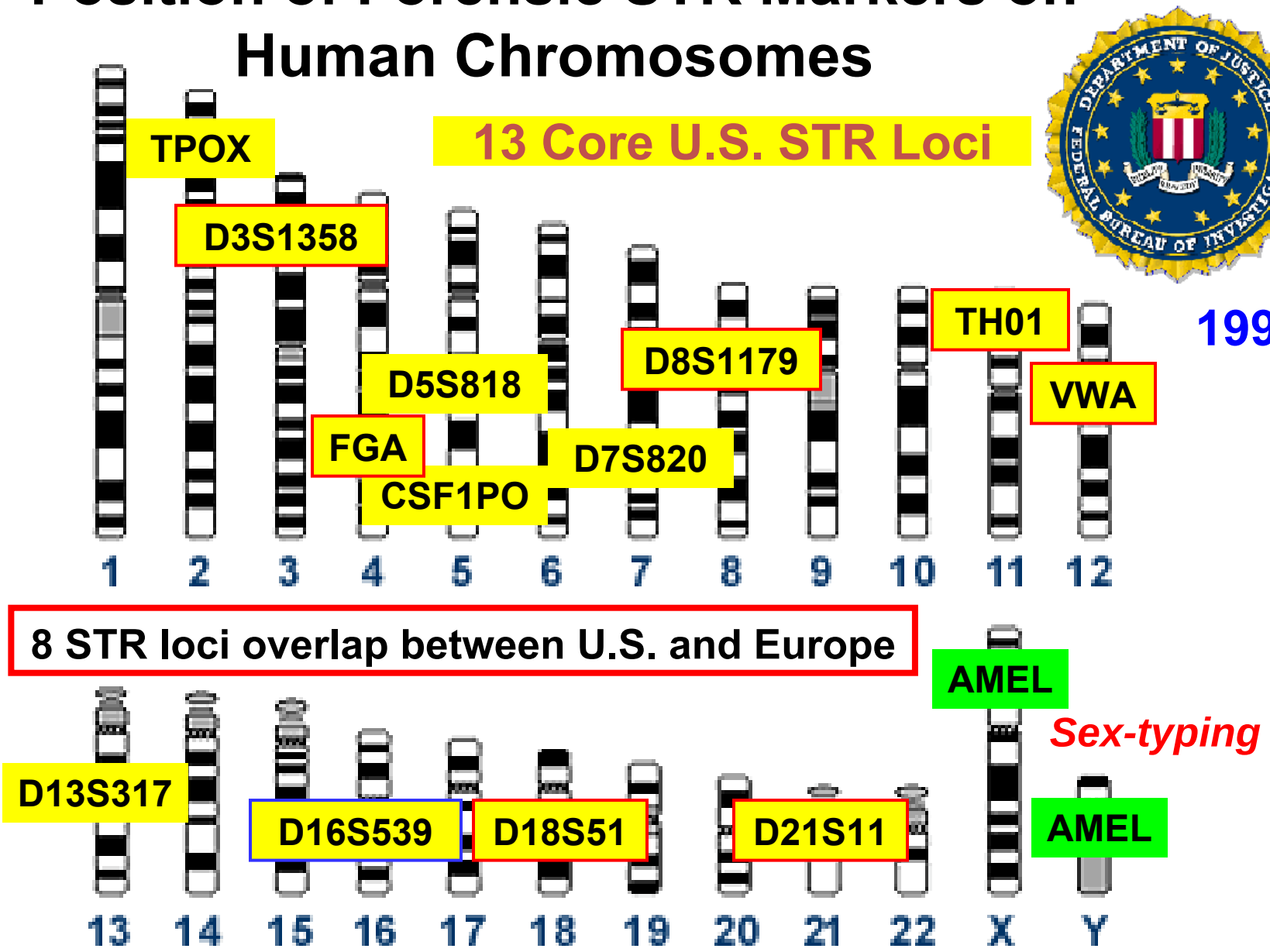
Individual #2



What would be entered into a DNA database for searching:

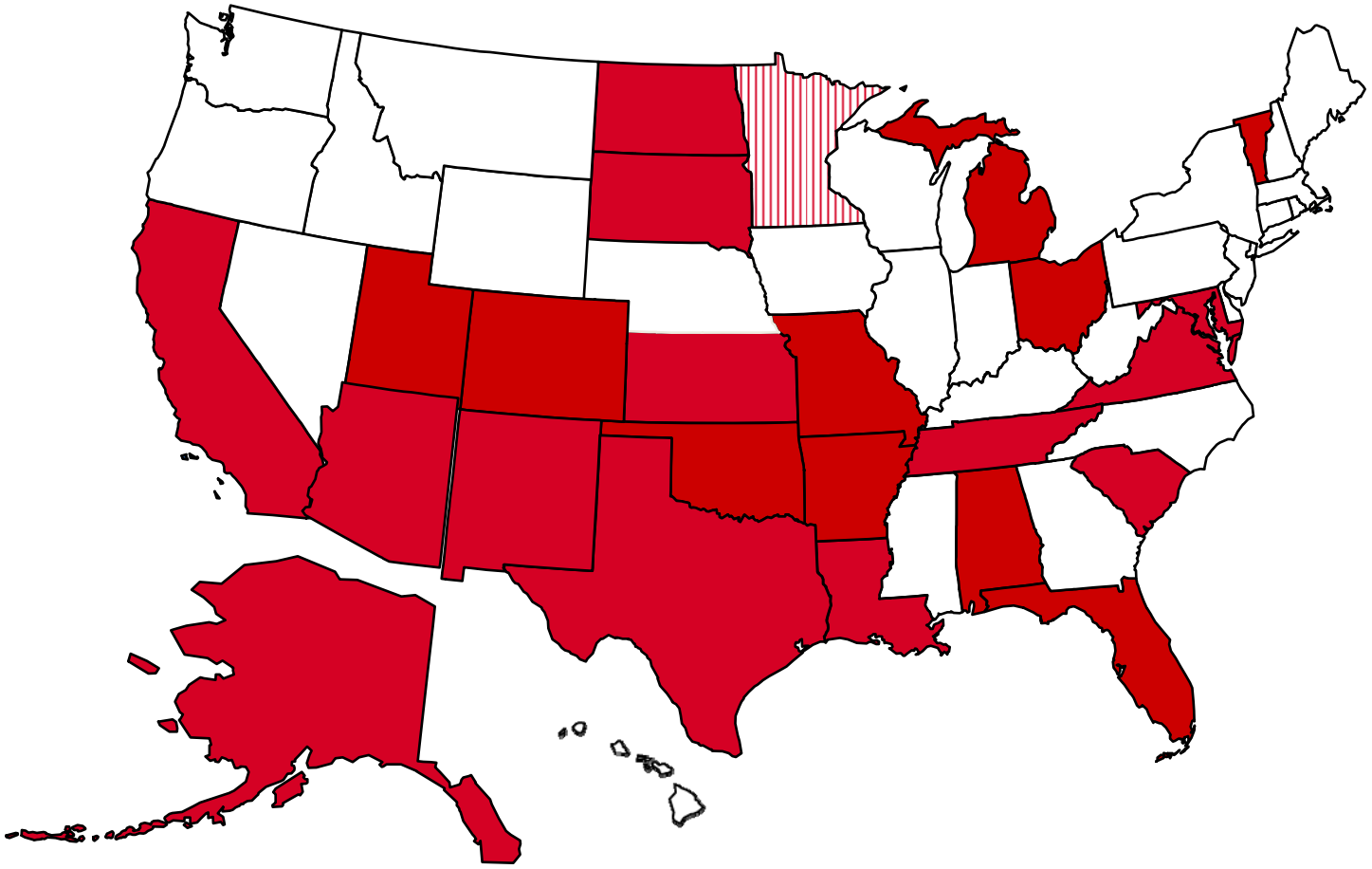
	AMEL	**D8S1179**	**D21S11**	**D18S51**
Individual #1	X,Y	11,13	28,32.2	17,18
Individual #2	X,X	11,14	30,31	12,15

Position of Forensic STR Markers on Human Chromosomes



Half of the U.S. Requires Arrestee DNA Testing

+ Federal & DoD



Data as of July 2010

Issues Facing DNA Databases

- Privacy Concerns with DNA Data
- Handling Technology Changes and Legacy Data
- Working Unknown Suspect Cases
- Sample Backlogs
- Sample Collection from Convicted Offenders
- Duplicate Samples or Twins
- Sample Retention
- Challenges with Sample/DNA Profile Expungement
- Measuring DNA Database Performance
- Follow-up to Database Matches

FBI Laboratory Backlog Mentioned in September 2007

DNA backlog piles up for FBI

Updated 9/3/2007 11:12 PM | Comments 34 | Recommend 30 E-mail | Save | Print | Re

By Richard Willing, USA TODAY

WASHINGTON — The FBI has fallen behind in processing DNA from nearly 200,000 convicted criminals — 85% of all samples it has collected since 2001 — Justice Department records show.

The backlog, which expands monthly, means most of the biological samples the bureau collects have not been stored in the national DNA database and used to solve crimes. DNA from 34,000 convicts has been added to the database since 2001, resulting in 600 matches to unsolved crimes, according to statistics furnished by the Justice Department to the Senate Judiciary Committee. At the same rate, the unloaded samples could help solve an additional 3,200 crimes.

Due to expanding collection laws
(often without supportive funding to do the work)

Backlog Elimination Schedule (2010 results)

- Progressive uploads of samples each month
  - January 15,000
  - February 18,000
  - March 25,000
  - April 35,000
  - May 45,000
  - June 65,000
  - July* 80,000
  - August* Balance (30,000)

Accomplished through adding automation to sample tracking, handling, and data interpretation

* The 145,000 sample upload across June and July resulted in over 1200 new hits

313,000 samples in 8 months

408,000 during FY2010 (Oct 2009 – Sept 2010)

Slide courtesy of FBI Laboratory Federal DNA Database Unit

Federal DNA Database Unit
37 FBI + 7 Contractors

- One Unit Chief/Technical Leader
- Three Supervisors
- Eight Examiners (5 qualified)
- Twenty Biologists (8 qualified)
- Four Management Program Analysts
- One Management Assistant
- Staffing Level of 37

- One Systems Integrator (Contractor)
- One Contractor Supervisor
- One Records Examiner (Contractor)
- Three Data Entry Clerks (Contractors)
- One Desk-Top Support Contractor

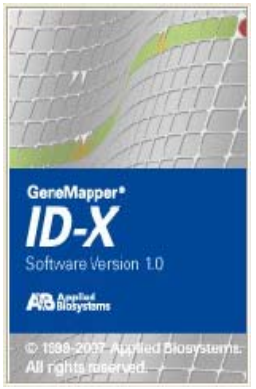
Slide courtesy of FBI Laboratory Federal DNA Database Unit

Phase III
High Throughput Automated DNA System

Semi-Automated sample prep

Robotic Sample Processing

High throughput Genetic Analyzer (3730)



Expert System Data Review

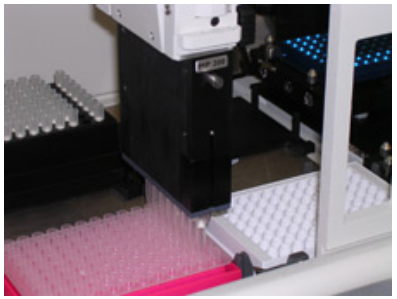
Slide courtesy of FBI Laboratory Federal DNA Database Unit

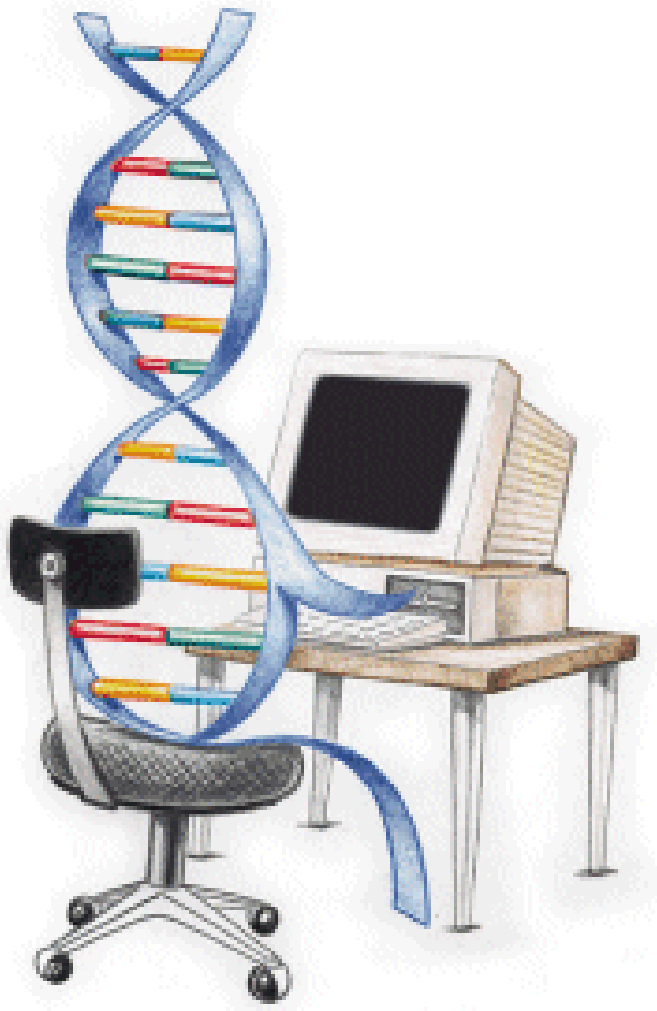
Laboratory Information Management System = STACS

- STACS = Sample Tracking and Control System (STaCS).
- Barcoding system (LIMS) that tracks all the information associated with samples, reagents, and instruments.

Slide courtesy of FBI Laboratory Federal DNA Database Unit

Future Predictions





- **More Automation**
- **Expert Systems**
- **Animal & Plant DNA**
- **Portable Devices**
- **Estimation of Physical Characteristics and Sample Ethnicity**

http://www.manastungare.com/publications/genetic/dna.gif

When There are No Hits After a DNA Database Search…

- **John Doe Warrants**
  - DNA profile from the evidence is filed as the offender to stop the clock on statute of limitations for commencing a criminal case
- **Mass Screens** (DNA Dragnets)
  - DNA samples are collected from a specific locality, age, gender, and often ethnic group to search for a matching profile to the crime scene evidence
- **Familial Searching**
  - The stringency of a search is reduced in order to look for a potential relative where DNA profile characteristics are shared with the evidence rather than a direct match

Biological Relatives Served as References

Captured December 13, 2003

Is this man really Sadaam Hussein?

Matching Y-STR Haplotype Used to Confirm Identity

(along with allele sharing from autosomal STRs)

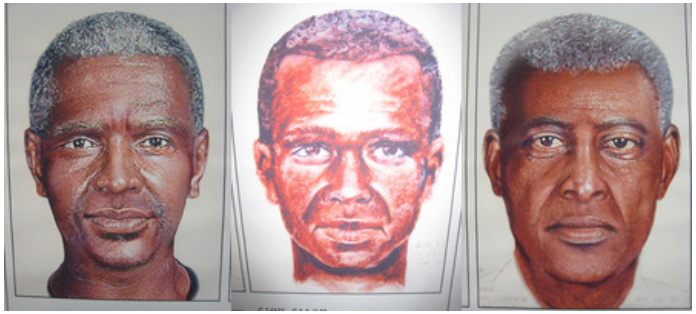
Uday and Qusay Hussein

Killed July 22, 2003

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Box 23.1, p. 534

L.A. Serial Killer Netted July 7, 2010 by a Familial DNA Search

Police sketches released in 2009



Arrested July 7, 2010

http://blogs.laweekly.com/informer/mugshotLonnie.png

Lonnie David Franklin Jr.

Franklin, a mechanic with a history of stealing cars, was arrested July 7 as he walked out of his mint green home on West 81st Street near Western Avenue after DNA evidence linked him to the crimes. **Franklin, 57, was caught through familial DNA testing after his son was arrested for a weapons charge in 2009 and had to give up a DNA swab.**

He is charged with 10 counts of murder and one count of attempted murder for a series of killings that date back to 1985.

http://blogs.laweekly.com/informer/crime/grim-sleeper-son-dna-trail-led/

California Familial DNA Search Team

Familial DNA Testing Scores A Win in Serial Killer Case

http://www.sciencemag.org/cgi/reprint/329/5989/262.pdf

Victims of the Grim Sleeper

http://www.laweekly.com/2008-08-28/news/eleven-lives-stolen-and-one-lucky-survivor/

The Grim Sleeper's Victims

1)Debra Jackson (age 29) – August 10, 1985
2)Henrietta Wright (age 35) – August 12, 1986
3)Thomas Steele (age 36) – August 14, 1986
4)Barbara Ware (age 23) – January 10, 1987
5)Bernita Sparks (age 25) – April 15, 1987
6)Mary Lowe (age 26) – October 31, 1987
7)Lachrica Jefferson (age 22) - January 30, 1988
8)Monique Alexander (age 18) – September 11, 1988
9)Enietra _____ (raped but survived) – November 1988

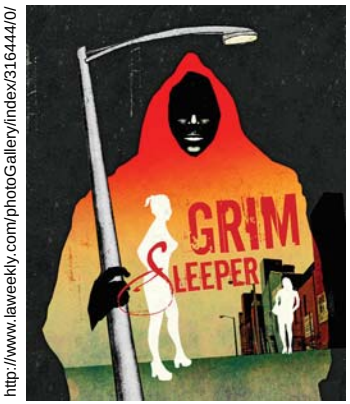
10) Princess Berthomieux (age 14) – March 19, 2002
11) Valerie McCorvey (age 35) – July 11, 2003
12) Janecia Peters (age 25) – January 1, 2007

http://blogs.laweekly.com/informer/crime/grim-sleeper-son-dna-trail-led/

Ballistics on bullets recovered from the victim's bodies matched

DNA evidence recovered

Over a 13 year gap in detected crimes, hence the "Sleeper" nickname



http://www.laweekly.com/photoGallery/index/316444/0/

Familial DNA Searching Performed with the Grim Sleeper Case

Lonnie David Franklin Jr., the man accused of being the Grim Sleeper serial killer, was caught in July 2010 when his son's DNA connected him to a series of crimes committed in L.A. over the past 25 years

"Nevertheless, familial DNA testing is an increasingly controversial technique. Critics such as the American Civil Liberties Union argue that familial DNA searches violate the Fourth Amendment prohibition against "unreasonable searches and seizures", as well as its "probable cause" clause. For instance, **should a possibly innocent relative be regularly "genetically surveilled" because their kinfolk has been in trouble with the law?**"

http://www.thegrio.com/opinion/how-familial-dna-can-help-crime-victims.php

National DNA Index System (NDIS)

No names are associated with DNA profiles uploaded to NDIS
If my profile was entered for searching:
16,17-17,18-21,22-12,14-28,30-14,16-12,13-11,14-9,9-9,11-6,6-8,8-10,10

FBI LABORATORY

http://www.fbi.gov/hq/lab/codis/index1.htm

Combined DNA Index System (CODIS)

Launched in October 1998 and now links all 50 states

Used for linking serial crimes and unsolved cases with repeat offenders

Convicted offender and forensic case samples along with a missing persons index

Requires 13 core STR markers

~130,000 investigations aided nationwide as of April 2011

Contains more than 9.5 million DNA profiles

NIST Applied Genetics

NIST Applied Genetics Group

Group Leader

John Butler

Marcia Holden

Margaret Kline

Pete Vallone

Dave Duewer*

Ross Haynes

Becky Hill

Erica Butts

Kristen O'Connor

Mike Coble

NIST Applied Genetics

APPLIED GENETICS Group

Major Programs Currently Underway

- **Forensic DNA**
  - New loci and assays (26plex)
  - STR kit testing
  - Ancestry SNP assays
  - Low-template DNA studies
  - Mixture interpretation
  - STR nomenclature
  - Variant allele cataloging and sequencing
  - Expert systems review
  - Training workshops to forensic DNA laboratories
  - Validation information and software tools
  - Textbook – 3rd ed. (3 volumes)
- **Clinical Genetics**
  - **Huntington's Disease SRM**
  - **CMV SRM**
  - **Exploring future needs**
- **Ag Biotech**
  - "universal" GMO detection/ quantitation (35S promoter)
- **DNA Biometrics**
  - Rapid PCR methods
  - Efforts to standardize testing of future portable DNA systems
  - Kinship analysis
- **Cell Line Authentication**

The Future of Forensic DNA

is Similar to the Olympic Motto of

"Swifter, Higher, Stronger"

Resources **Training** **Action**

Recent NIST Publications Demonstrating "Swifter, Higher, Stronger" DNA Analysis

Swifter PCR Amplification

Forensic Science International: Genetics Supplement Series 2 (2009) 111–112

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

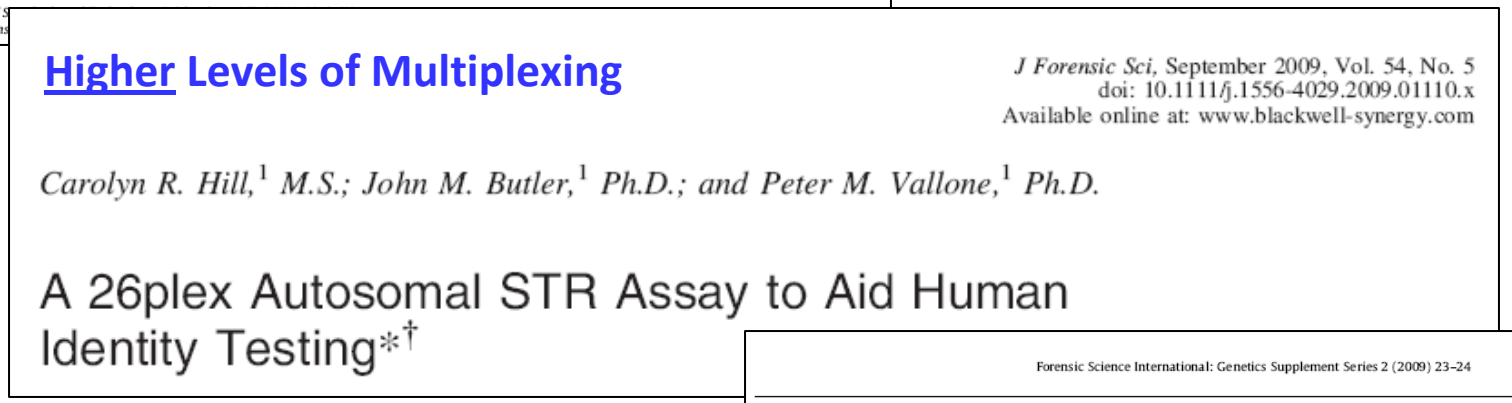
Research article

Rapid amplification of commercial STR typing kits

Peter M. Vallone[a,*], Carolyn R. Hill[a], Daniele Podini[b], John M. Butler[a]

[a] National Institute of S
[b] Department of Forens

Higher Levels of Multiplexing

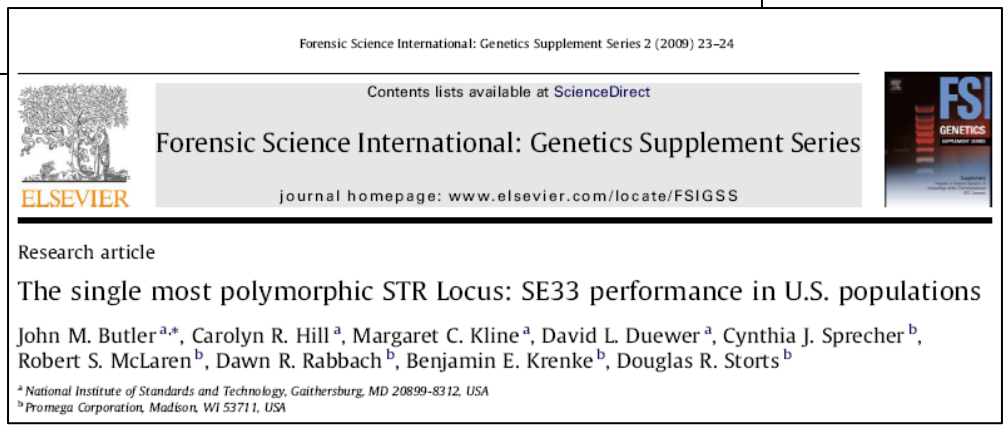


J Forensic Sci, September 2009, Vol. 54, No. 5
doi: 10.1111/j.1556-4029.2009.01110.x
Available online at: www.blackwell-synergy.com

Carolyn R. Hill,[1] *M.S.; John M. Butler,*[1] *Ph.D.; and Peter M. Vallone,*[1] *Ph.D.*

A 26plex Autosomal STR Assay to Aid Human Identity Testing*†

Stronger Powers of Discrimination



Forensic Science International: Genetics Supplement Series 2 (2009) 23–24

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Research article

The single most polymorphic STR Locus: SE33 performance in U.S. populations

John M. Butler[a,*], Carolyn R. Hill[a], Margaret C. Kline[a], David L. Duewer[a], Cynthia J. Sprecher[b], Robert S. McLaren[b], Dawn R. Rabbach[b], Benjamin E. Krenke[b], Douglas R. Storts[b]

[a] National Institute of Standards and Technology, Gaithersburg, MD 20899-8312, USA
[b] Promega Corporation, Madison, WI 53711, USA

Acknowledgments

- FBI Laboratory's Federal DNA Database Unit
  - For sharing their recent experience with automation to improve sample throughput

- Funding support from the National Institute of Justice and the FBI

- Support from an excellent group of research scientists within the NIST Applied Genetics Group

Thank you for your attention

Contact Information

John Butler
NIST Fellow
Group Leader of Applied Genetics
john.butler@nist.gov
301-975-4049

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

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