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.

268 exonerated as of April 19, 2011
Historical Perspective on DNA Typing

1985: PCR developed

1990: RFLP

1992: Capillary electrophoresis of STRs first described

1994: First STRs developed

1996: CODIS loci defined


1998: FSS Quadruplex

1998: DQA1 & PM (dot blot)

2000: STR typing with CE is fairly routine

2002: PowerPlex 16 (16 loci in single amp)

2004: NDIS launched (October 13, 1998)

2004: CODIS loci defined

2011: DNA is an important part of the criminal justice system

www.dna.gov
President’s DNA Initiative
Debbie Smith Act Backlog Reduction (>1B from 2004-2010)

2011: Identifiler 5-dye kit and ABI 3100

2011: miniSTRs

2011: Y-STRs

2011: mtDNA

2011: Multiplex STRs

The DNA Field Moves Forward…

The Past

Sexual Assault Case

- Known
- Evidence
- Known

LADDER Control
LADDER Victim
Suspect 1
Female Cells
DNA

The Present

RFLP

STRs

The Future
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

• **Exclusion** (no match)

• **Non-exclusion**
  – “Match” or “inclusion”

• **Inconclusive result**
  (or a complex mixture)
DNA Profile Comparison

**DNA Profile Comparison**

1. **Exclusion (no match)**
   - **Q ≠ K**
   - May be inconclusive due to Forensic Issues (degradation, mixtures, low levels)

2. **Inclusion (match)**
   - **Q = K**
   - May be inconclusive due to Lack of Available Reference

**Steps Involved**

- **Collection**
- **Sample Storage**
- **Extraction**
- **Quantitation**
- **Amplification**
- **STR Markers**
- **Separation/Detection**
- **Data Interpretation**
- **Statistical Interpretation**

**Suspect developed**

- **Reference (Known) sample “K”**
- **Profile put on database**

**Evidence (Question) sample “Q”**

- **Crime committed**
  - Biological material transferred

**Report (with statistical weight)**

- **Plea**
- **Court**

**Profile put on database**

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

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

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

For most recent data, see [http://ag.ca.gov/bfs/pdf/Monthly.pdf](http://ag.ca.gov/bfs/pdf/Monthly.pdf)
Advocates for DNA Funding and Expansion

Debbie Smith

Mitch Morrissey

Kirk Bloodsworth

Victim
Prosecutor
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

Growth due in part to federal funding from the Debbie Smith Act and new DNA collection laws

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)

**Biology**
- Blood Stain
- Buccal swab
- Sample Collection & Storage
- DNA Extraction
- DNA Quantitation
- Multiplex PCR Amplification
- STR Typing
- DNA separation and sizing
- Interpretation of Results

**Genetics**
- Statistics Calculated
- DNA Database search
- Paternity test
- Reference sample
- Applied Use of Information

**Technology**
Short Tandem Repeat (STR) Markers

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

The overall PCR product size is measured

The PCR product size is generated

Fluorescent dye

Forward PCR primer

Reverse PCR primer

DNA template containing STR marker

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

The overall PCR product size is measured

PCR Product Size (bp)

Allelic Ladder

Sample #1

Sample #2

TCCCAAGCTCTTCCCTTCCCTCCTTCCCTAGATCAATACAGACAGAGACAGGTGGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAGATAGATAGATACATGCTTACAGATGCACAC

= 11 GATA repeats (“11” is all that is reported)
Short Tandem Repeat (STR) Typing

Fluorescent dye-labeled primer

STR Repeat Region

(formal)

(formal)

forward primer hybridization region

reverse primer hybridization region

(size in bp)

75….80….100….120….140….160….180….200….220….240….260…..

RFUs

DNA Separation and Detection

1000

500

139bp

147bp
Family Inheritance of STR Alleles (D13S317)

**PCR product size (bp)**

<table>
<thead>
<tr>
<th></th>
<th>180</th>
<th>190</th>
<th>200</th>
<th>210</th>
<th>220</th>
<th>230</th>
<th>240</th>
<th>250</th>
</tr>
</thead>
</table>
| Father | ![Father Peaks](chart)
| Child #1 | ![Child #1 Peaks](chart)
| Child #2 | ![Child #2 Peaks](chart)
| Child #3 | ![Child #3 Peaks](chart)
| Mother | ![Mother Peaks](chart)

**PATERNITY TESTING**
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

What would be entered into a DNA database for searching:

<table>
<thead>
<tr>
<th></th>
<th>AMEL</th>
<th>D8S1179</th>
<th>D21S11</th>
<th>D18S51</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual #1</td>
<td>X,Y</td>
<td>11,13</td>
<td>28,32.2</td>
<td>17,18</td>
</tr>
<tr>
<td>Individual #2</td>
<td>X,X</td>
<td>11,14</td>
<td>30,31</td>
<td>12,15</td>
</tr>
</tbody>
</table>
Position of Forensic STR Markers on Human Chromosomes

13 Core U.S. STR Loci

- TPOX
- D3S1358
- D5S818
- FGA
- CSF1PO
- D8S1179
- D7S820
- TH01
- VWA

8 STR loci overlap between U.S. and Europe

- D13S317
- D16S539
- D18S51
- D21S11

1997

Sex-typing

AMEL

AMEL
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
DNA backlog piles up for FBI

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)

* 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.
Future Predictions

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

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

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.

Familial DNA Testing Scores A Win in Serial Killer Case

Victims of the Grim Sleeper


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

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

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?”

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 Group

Group Leader

John Butler
Marcia Holden
Margaret Kline
Pete Vallone
Dave Duewer*
Ross Haynes
Becky Hill
Erica Butts
Kristen O’Connor
Mike Coble
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”
Recent NIST Publications Demonstrating “Swifter, Higher, Stronger” DNA Analysis

Swifter PCR Amplification


Contents lists available at ScienceDirect

Research article
Rapid amplification of commercial STR typing kits

Peter M. Vallone*, Carolyn R. Hill*, Daniele Podini*, John M. Butler*  

Higher Levels of Multiplexing

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


Contents lists available at ScienceDirect

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


* National Institute of Standards and Technology, Gaithersburg, MD 20899-8312, USA  
† Permgga Corporation, Madison, WI 53711, USA  
‡ Permgga 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