Forensic DNA Typing and Prospects for Biometrics

Dr. John M. Butler
NIST Biotechnology Division
May 12, 2004

Examples of DNA in the News
- Saddam Hussein Identification
- Source of Cow with "Mad cow" Disease
- Scot Peterson Murder Trial
- Identification of WTC Victims
- "Thomas Jefferson fathered slave’s children"

Examples of DNA in the News
- Saddam Hussein Identification
- Source of Cow with "Mad cow" Disease
- Scot Peterson Murder Trial
- Identification of WTC Victims
- "Thomas Jefferson fathered slave’s children"

DNA Profile

Decoded

Saddam Hussein’s capture verified with DNA testing

Source: www.cnn.com; The Scientist Dec 19, 2003

Portion of DNA Profile

Captured December 13, 2003

"A Forensic Paternity Test"

DNA Used as the Ultimate Biometric for Confirming Saddam’s Identity

Our DNA Comes from our Parents

Father’s Sperm

Mother’s Egg

Child’s Cell

Genetic Inheritance Pattern of DNA Profiles

DAD

MOM

CHILD

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

**Family Inheritance of STR Alleles (D13S317)**

<table>
<thead>
<tr>
<th>PCR product size (bp)</th>
<th>12</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child #1</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Child #2</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Child #3</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Mother</td>
<td>8</td>
<td>12</td>
</tr>
</tbody>
</table>

#### Results of DNA Tests Impact Families

- Amanda
- Marshall
- Katy
- Me
- My Wife

### Historical Investigation DNA Study

(Matching Relatives to Remains or Relatives to Relatives)

- Thomas Woodson
- Different Y Haplotype

- Thomas Jefferson II
- Field Jefferson
- Same Y Haplotype

- President Thomas Jefferson
- Father Estonia Hemings

### Methods for Human Identification

- Fingerprints have been used since 1901
- DNA since 1986

### Forensic DNA Testing

The genome of each individual is unique (with the exception of identical twins)

- Probe subsets of genetic variation in order to differentiate between individuals

- DNA typing must be done efficiently and reproducibly (information must hold up in court)

- Typically, we are not looking at genes – little/no information about race, predisposition to disease, or phenotypical information (eye color, height, hair color) is obtained
Applications for Human Identity Testing

- Forensic cases - matching suspect with evidence
- Paternity testing - identifying father
- Historical investigations
- Missing persons investigations
- Mass disasters - putting pieces back together
- Military DNA “dog tag”
- Convicted felon DNA databases

As DNA analysis has shown its usefulness, the number of samples gathered for testing purposes has gone up dramatically...

Roles of Biological Evidence in Criminal Investigation

- Identify a person
- Exclude a suspect
- Link suspect, victim and crime scene
- Link weapon to victim
- Link witness to scene
- Prove or disprove an alibi
- Reconstruct the scene
- Provide investigative leads

Sources of Biological Evidence

- Blood
- Semen
- Saliva
- Urine
- Hair
- Teeth
- Bone
- Tissue

Blood stain

Only a very small amount of blood is needed to obtain a DNA profile

Steps in DNA Analysis

- Collection
- Specimen Storage
- Extraction
- Quantitation
- Genotyping
- Interpretation of Results
- Database Storage & Searching

Blood Stain

Species

DNA Extraction

Multiplex PCR Amplification

STR Typing

Main: 13, 14-15, 18-12, 13-10, 13-15, 16

Interpretation of Results

DNA Database

Armed Forces DNA Repository

- Located in Gaithersburg, Maryland

>4.5 million blood cards on file from members of U.S. military

Post-conviction DNA testing is used to exonerate falsely incarcerated individuals (143 so far)
DNA in the Cell

- DNA molecule in the cell nucleus
- Double-stranded DNA
- ~3 billion total base pairs
- Target Region for PCR

What Type of Genetic Variation?
- Length Variation
  - Short tandem repeats (STRs)
    - CTAGTCGT(GATA)(GATA)(GATA)CGGATCGT
- Sequence Variation
  - Single nucleotide polymorphisms (SNPs)
  - Insertions/deletions
    - GCTAGTCGATGCTC(G/A)CGTATGCTGTAGC

Basic Concepts

- PCR polymerase chain reaction: a method of amplifying a specific region of the genome
- Locus: region of the genome being examined
- Allele: the state of the genetic variation being examined
  - STRs = number of repeat units
  - SNPs = base sequence at the site
- Chromosomes are paired so...
  - Homozygous: alleles are identical on each chromosome
  - Heterozygous: alleles differ on each chromosome

Short Tandem Repeats (STRs)

- Fluorescent dye creates a labeled PCR product
- Primer positions define PCR product size
- Primer positions define PCR product size
- The repeat region is variable between samples while the flanking regions where PCR primers bind are constant

Position of Forensic STR Markers on Human Chromosomes

- 13 CODIS Core STR Loci
- Primer positions define PCR product size
- Primer positions define PCR product size
- Primer positions define PCR product size
- Primer positions define PCR product size

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

Inlet Buffer → Capillary filled with polymer solution → Outlet Buffer

Sample tray moves automatically beneath the cathode end of the capillary to deliver each sample in succession.

Data Acquisition

Sample Detection

Butler, J.M. (2001) Forensic DNA Typing, Figure 9.3, ©Academic Press

Principles of Sample Separation and Detection

Labeled DNA fragments (PCR products)

Size Separation

Color Separation

CCD Panel

Fluorescence

ABI Prism spectrograph

Principles of Sample Separation and Detection

Scanned Gel Image

Capillary Electropherogram

Butler, J.M. (2001) Forensic DNA Typing, Figure 10.8, ©Academic Press

Capillary Electrophoresis Instrumentation

ABI 310 single capillary

ABI 3100 16-capillary array

 Allelic Ladders

8

11

14

All heterozygous alleles

PCR Product Size (bp)

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

For heterozygous loci

\[ P = 2pq \]

\( P \) = probability; \( p \) and \( q \) are frequencies of allele in a given population

Example: For the locus D3S1358 and individual is 16,17 with frequencies of 0.2315 and 0.2118 respectively

\[ P = 2(0.2315)(0.2118) = 0.0981 \text{ or } 1 \text{ in } 10.2 \]

For independent loci, the genotype frequencies can be combined through multiplication...

Profile Probability = \( (P_1)(P_2) \ldots (P_n) \)

= 1 in a very large number...

<table>
<thead>
<tr>
<th>Locus</th>
<th>D3S1358</th>
<th>D5S818</th>
<th>D7S820</th>
<th>D13S317</th>
<th>D16S539</th>
<th>D18S51</th>
<th>D19S433</th>
<th>TH01</th>
<th>TPOX</th>
<th>CSF1PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>16</td>
<td>12</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Value</td>
<td>0.2315</td>
<td>0.1459</td>
<td>0.1459</td>
<td>0.1459</td>
<td>0.1459</td>
<td>0.1459</td>
<td>0.1459</td>
<td>0.1459</td>
<td>0.1459</td>
<td>0.1459</td>
</tr>
<tr>
<td>Value</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
<td>0.2118</td>
</tr>
<tr>
<td>Value</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
</tr>
<tr>
<td>Value</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
</tr>
<tr>
<td>Value</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
</tr>
<tr>
<td>Value</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
</tr>
<tr>
<td>Value</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
</tr>
<tr>
<td>Value</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
<td>1024</td>
</tr>
</tbody>
</table>

| What would be entered into a DNA database for searching |
|-------|--------|--------|--------|
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |
| 16,17 | 16,17  | 16,17  |

The Random Match Probability for this profile in the FBI Caucasian population is 1 in 1.56 quadrillion (10^{15})
**CODIS DNA Database**

**Combined DNA Index System**
- Used for linking serial crimes and unsolved cases with repeat offenders
- Convicted offender and forensic case samples
- Launched October 1998 and links all 50 states
- Requires 13 core STR markers
- Current backlog of >750,000 samples (millions of dollars are spent each year to reduce backlog)

**National DNA Database**
- 1,641,076 offenders
- 78,475 forensic
- 102 missing persons
- 175 CODIS labs in 50 states, FBI, US Army Crime Lab & Puerto Rico
- 16,160 investigations aided

**Results from Virginia**

<table>
<thead>
<tr>
<th>Statistical Information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offender Profiles</td>
<td>185,387</td>
</tr>
<tr>
<td>Forensic Samples</td>
<td>3,068</td>
</tr>
<tr>
<td>Number of CODIS Labs</td>
<td>4</td>
</tr>
<tr>
<td>NDIS Participating Labs</td>
<td>4</td>
</tr>
<tr>
<td>Investigations Aided</td>
<td>1,773</td>
</tr>
</tbody>
</table>

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

**Ensuring Accurate Forensic DNA Results**

- ASCLD-LAB Accreditation
- Proficiency Testing of Analysts
- DAB Standards
- SWGDAM Guidelines
- NIST Standard (SRMs)
- Inspections/ Audits

**STRBase**

- Short Tandem Repeat DNA Internet Database
- Prepares Standard Reference Materials (SRMs)
  - SRM 2391b PCR-based DNA Profiling Standard
  - SRM 2395 Human Y-Chromosome DNA Profiling Standard
- Creates databases with useful information
  - STRBase (http://www.cstl.nist.gov/biotech/strbase)
  - Evaluates and develops new technologies
  - Conducts interlaboratory testing
  - Performs quality control testing for labs & companies

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

**Results from Virginia**

- 1,773 investigations aided
- 4 NDIS Participating Labs
- 3,068 Forensic Samples
- 185,387 Total Samples
- 1,641,076 Total Offenders
- 78,475 Total Forensic
- 102 Total Missing Persons
- 175 CODIS Labs
- 16,160 Investigations Aided

**General Information**
- Intro to STRs (downloadable PowerPoint)
- STR Fact Sheets
- Sequence Information
- Multiplex STR Kits
- Variant Allele Reports

**Forensic Interest Data**
- FBI CODIS Core Loci
- DAB Standards
- NIST SRM 2391
- Published PCR Primers
- Y-Chromosome STRs
- Population Data
- Validation Studies

**Supplemental Info**
- Reference List
- Technology Review
- Addresses for Scientists
- Links to Other Web Sites

http://www.cstl.nist.gov/biotech/strbase
New DNA Test for Cats Developed in Our Lab
(the “MeowPlex”)

Different peak patterns between samples

Male
Female

Telling Cats Apart…

“Kitty Biometrics”

Disclaimer

- Our laboratory performs research and development in human identification techniques involving DNA testing
- I am not an “expert” in biometrics and therefore do not understand all of the issues involved
- The following ideas are my own and in no way represent the official position of the National Institute of Standards and Technology (NIST)

DNA within the Biometric Model

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

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

Issues with DNA as a Biometric Tool

Individualization of everyone except identical twins
- Invasive collection of test sample (swab of cheek cells or finger prick for blood)
- Concern over genetic privacy if sample is stored (would or could sample ever be used for other purposes?)
- Expense for testing (> $20 per sample; forensic testing is > $500)
- Time required for testing (minimum 4-5 hours with forensic markers)
- Database search/storage of DNA profile (who controls access?; would it ever be used for criminal investigations?)
- Development of reference databases for terrorists, etc. (how would a match be used?)
- Chance for falsifying your sample (e.g., movie GATTACA)

Collection of Test Sample

- Buccal (cheek cell) swab is not as invasive as a finger prick to draw blood
- DNA can be reliably extracted off of the paper or Q-tip swab

Used routinely in Virginia for felon arrestee DNA collection with 99.5% success (only 23 failures in first 6 months of use)
Genetic Privacy Concerns

- The DNA profile itself is neutral (and uninformative)—just a string of numbers like a Social Security Number.
- DNA markers used in forensics were selected to be neutral and are located away from or between genes rather than being part of gene products and therefore are not generally thought to be associated with any genetic disease.
- Concern is really with the DNA sample collected—would it be retained and used for any other type of testing? (e.g., Armed Forces DNA Repository can only be used for identifying combat casualties).

Time Required for Testing

| Collection | Could be <5 minutes |
| Extraction | Not necessary if samples are uniform in amount |
| Quantitation | Rapid thermal cycling to be done with microwave, typically 2-3 hours |
| Amplification | DNA separations (STR analysis) of <5 minutes have been demonstrated; typically ~30 minutes |
| Genotyping | Currently performed manually in most labs; expert systems are under development to enable rapid interpretation |
| Interpretation of Results | Search could be similar to fingerprint search in terms of speed |

Comparison a DNA profile to a reference or database

Male: 13,14-15,16-12,13-10,13-15,16-…..

Possible Ways DNA Could Be Used Now

As part of check before issuing visa for foreign visitors wanting to visit U.S.:  
- DNA samples could be shipped from each embassy to U.S.-based lab for testing.
- Visa would not be issued until DNA profile was generated and searched against a national/international database.
- DNA profile would be included in biometric passport for future confirmation of identity as needed.

Confirming identity upon entry into the U.S. while the plane is in the air:
- Would require DNA labs operating in foreign airports.
- 200+ DNA profiles would be generated in ~8 hours (before a plane coming from Europe landed).

In FY2003, 7,300,667 U.S. passports were issued.

How close are we to GATTACA?

Gattaca Corp. is an aerospace firm in the future. During this time society analyzes your DNA and determines where you belong in life. Ethan Hawke's character was born with a congenital heart condition which would cast him out of getting a chance to travel in space. So in turn he assumes the identity of an athlete named Jerome who has genes that would allow him to achieve his dream of space travel. Hawke's character Vincent constantly passes genetic tests by diligently using samples of Jerome's hair, skin, blood and urine. Entry to secure locations controlled by rapid genetic profiling.

Ability to predict future health risks based on genetic testing.

How close are we to GATTACA?  

NOT THERE YET!

Sex Determination PCR from human buccal cells

Multiplex PCR (dye) from human buccal cells without DNA extraction

20 cycles PCR after hybridization

15 minutes for PCR amplification and detection

Lagally et al., Lab-on-a-Chip, 1, 102 (2001)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Virginia DNA Testing of Felon Arrestees

As of January 1, 2003, any individual arrested for a violent felony crime (Code of Virginia § 19.2-310.2-1) must provide a buccal sample for DNA analysis, with the resultant profile incorporated into the Virginia DNA Data Bank (Code of Virginia § 19.2-310.5).

Since January 2003
- Buccal swab collected upon arrest
- DNA sample processed within 72 hours
- DNA profile searched against state database (national database does not currently allow searches for individuals prior to conviction)
- If a match results, then arrestee is detained and later prosecuted
- From Jan 2003 – Dec 2003, VA processed 7,836 arrestee samples (not all analyzed) and scored 63 hits against their state database (Profiles in DNA, 2004, 7(1):3-5)

If you want to know more…
- Forensic DNA Typing: Biology and Technology behind STR Markers
- NIST website: http://www.cstl.nist.gov/biotech/strbase
- John Butler email: john.butler@nist.gov

Thank you for your attention…

NIST Project Team:
- John Butler (leader)
- Margaret Kline
- Jan Redman
- Pete Vallone
- Dave Duane
- Jeff Agleby
- Amy Decker
- Mike Coble

National Institute of Justice
Funding through NIST Office of Law Enforcement Standards

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