A Primer on DNA Profiling Using STR Markers

Dr. John M. Butler

National Institute of Standards and Technology

Biotechnology Division
Presentation Outline

- Biology and technology behind short tandem repeat (STR) DNA testing
- How statistical calculations are made with STRs
- Approaches for “challenging” samples: perspectives for the future
- Resources for additional information
Human Identity Testing

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

Involves generation of DNA profiles usually with the same core STR (short tandem repeat) markers
Basis of DNA Profiling

The genome of each individual is unique (with the exception of identical twins) and is inherited from parents.

Probe subsets of genetic variation in order to differentiate between individuals (statistical probabilities of a random match are used).

DNA typing must be performed efficiently and reproducibly (information must hold up in court).

Current standard DNA tests DO NOT look at genes – little/no information about race, predisposition to disease, or phenotypical information (eye color, height, hair color) is obtained.
Short Tandem Repeat (STR) Markers

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

**PCR product size generated**

**DNA template containing STR marker**

**Fluorescent dye**

**Forward PCR primer**

**Reverse PCR primer**

**PCR Product Size (bp)**

**Allelic Ladder**

Sample #1

Sample #2

TCCCAAGCTCTTCTTCTTTCCCTCCTAGATCAATACAGACAGA
AGACAGCTTGAGATAGATAGATAGATAGATAGATAGATAGATA
GATAGATAGATAGATAGATAGATACATGCTTACAGATGACAC

= 11 GATA repeats ("11" is all that is reported)
Advantages for STR Markers

- Use of the polymerase chain reaction (PCR) enables recovery of information from small amounts of material

- Small product sizes are generally compatible with degraded DNA

- Multiplex amplification with fluorescence detection enables high power of discrimination in a single test

- Commercially available in an easy to use kit format

- Uniform set of core STR loci provide capability for national and international sharing of criminal DNA profiles
Position of Forensic STR Markers on Human Chromosomes

13 CODIS Core STR Loci

- TPOX
- D3S1358
- D5S818
- FGA
- CSF1PO
- D8S1179
- D7S820
- D13S317
- D16S539
- D18S51
- D21S11
- TH01
- VWA
- AMEL

Sex-typing

Core STR Loci for the United States

1997
Position of Forensic STR Markers on Human Chromosomes

10 SGM Plus Loci
SE33 (Germany)

1995
1999

Core STR Loci for Europe

Sex-typing
Typical Instruments Used for STR Typing

Thermal Cycler for PCR Amplification

Capillary electrophoresis instruments for separating and sizing PCR products

**single capillary**
- ABI 310

**16-capillary array**
- ABI 3100

GeneAmp 9700
Steps in STR Typing

Sample Injection

- Mixture of dye-labeled PCR products from multiplex PCR reaction

Size Separation

- Sample Separation

Fluorescence

Color Separation

Capillary (filled with polymer solution)

Sample Detection

- ABI Prism spectrograph
- CCD Panel (with virtual filters)

Sample Interpretation

Processing with GeneScan/Genotyper software

FMBIO III Gel Imager System

Gel Electrophoresis

Gel Scanner

Gel Image

PowerPlex 16 BIO
Steps in DNA Analysis

**Collection**
- Blood Stain
- Buccal swab

**Specimen Storage**

**Extraction**
- Sample Collection & Storage
- DNA Extraction
- DNA Quantitation

**Quantitation**

**Genotyping**

**Interpretation of Results**

**Database**
- Storage & Searching

---

**Usually 1-2 day process (a minimum of ~5 hours)**

**Sample Collection & Storage**

**DNA Extraction**

**DNA Quantitation**

**Multiplex PCR**
- (Amplification of STR Loci)

**STR Typing**
- (DNA separation)

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

**Interpretation of Results**

**DNA Database**
Commercial STR 16plex Kits

**Identifiler™ kit** (Applied Biosystems)
multiplex STR result

13 core STR loci + 2 additional loci + AMEL sex-typing

**PowerPlex® 16 kit** (Promega Corporation)
multiplex STR result

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology
How Statistical Calculations are Made

- **Generate data** with set(s) of samples from desired population group(s)
  - Generally only 100-150 samples are needed to obtain reliable allele frequency estimates

- **Determine allele frequencies** at each locus
  - Count number of each allele seen

- Allele frequency information is used to **estimate the rarity** of a particular DNA profile
  - Homozygotes ($p^2$), Heterozygotes ($2pq$)
  - Product rule used (multiply locus frequency estimates)

For more information, see Chapters 20 and 21 in *Forensic DNA Typing, 2nd Edition*
The **11,12** genotype was seen **54 times** in 302 samples (604 examined chromosomes).
## Allele Frequency Tables

<table>
<thead>
<tr>
<th>Allele</th>
<th>Caucasian N=302</th>
<th>Caucasian N=7,636</th>
<th>African American N=258</th>
<th>African American N=7,602</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>0.0017*</td>
<td>0.0009</td>
<td>11</td>
<td>0.0003*</td>
</tr>
<tr>
<td>12</td>
<td>0.0017*</td>
<td>0.0007</td>
<td>12</td>
<td>0.0045</td>
</tr>
<tr>
<td>13</td>
<td>--</td>
<td>0.0031</td>
<td>13</td>
<td>0.0019*</td>
</tr>
<tr>
<td>14</td>
<td>0.1027</td>
<td>0.1240</td>
<td>14</td>
<td>0.0892</td>
</tr>
<tr>
<td>15</td>
<td>0.2616</td>
<td>0.2690</td>
<td>15</td>
<td>0.3023</td>
</tr>
<tr>
<td>15.2</td>
<td>--</td>
<td>--</td>
<td>15.2</td>
<td>0.0019*</td>
</tr>
<tr>
<td>16</td>
<td>0.2533</td>
<td>0.2430</td>
<td>16</td>
<td>0.3353</td>
</tr>
<tr>
<td>17</td>
<td>0.2152</td>
<td>0.2000</td>
<td>17</td>
<td>0.2054</td>
</tr>
<tr>
<td>18</td>
<td>0.15232</td>
<td>0.1460</td>
<td>18</td>
<td>0.0601</td>
</tr>
<tr>
<td>19</td>
<td>0.01160</td>
<td>0.0125</td>
<td>19</td>
<td>0.0039*</td>
</tr>
<tr>
<td>20</td>
<td>0.0017*</td>
<td>0.0001*</td>
<td>20</td>
<td>0.0048</td>
</tr>
</tbody>
</table>

The most common allele in the Caucasian population is allele 15 with a frequency of 0.2616. In the African American population, allele 15.2 has the highest frequency at 0.0019. Allele frequencies denoted with an asterisk (*) are below the 5/2N minimum allele threshold recommended by the National Research Council report (NRCII) *The Evaluation of Forensic DNA Evidence* published in 1996.

Einum et al. (2004) *JFS* 49(6)
### DNA Profile Frequency with all 13 CODIS STR loci

<table>
<thead>
<tr>
<th>Locus</th>
<th>allele</th>
<th>allele</th>
<th>value</th>
<th>value</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>16</td>
<td>17</td>
<td>0.2533</td>
<td>0.2152</td>
<td>9.17</td>
</tr>
<tr>
<td>VWA</td>
<td>17</td>
<td>18</td>
<td>0.2815</td>
<td>0.2003</td>
<td>8.87</td>
</tr>
<tr>
<td>FGA</td>
<td>21</td>
<td>22</td>
<td>0.1854</td>
<td>0.2185</td>
<td>12.35</td>
</tr>
<tr>
<td>D8S1179</td>
<td>12</td>
<td>14</td>
<td>0.1854</td>
<td>0.1656</td>
<td>16.29</td>
</tr>
<tr>
<td>D21S11</td>
<td>28</td>
<td>30</td>
<td>0.1589</td>
<td>0.2782</td>
<td>11.31</td>
</tr>
<tr>
<td>D18S51</td>
<td>14</td>
<td>16</td>
<td>0.1374</td>
<td>0.1391</td>
<td>26.18</td>
</tr>
<tr>
<td>D5S818</td>
<td>12</td>
<td>13</td>
<td>0.3841</td>
<td>0.1407</td>
<td>9.25</td>
</tr>
<tr>
<td>D13S317</td>
<td>11</td>
<td>14</td>
<td>0.3394</td>
<td>0.0480</td>
<td>30.69</td>
</tr>
<tr>
<td>D7S820</td>
<td>9</td>
<td>11</td>
<td>0.1772</td>
<td>0.3212</td>
<td>13.8</td>
</tr>
<tr>
<td>D16S539</td>
<td>9</td>
<td>11</td>
<td>0.1126</td>
<td>0.3212</td>
<td>13.8</td>
</tr>
<tr>
<td>THO1</td>
<td>6</td>
<td>11</td>
<td>0.2318</td>
<td>0.3212</td>
<td>13.8</td>
</tr>
<tr>
<td>TPOX</td>
<td>8</td>
<td>11</td>
<td>0.5348</td>
<td>0.3212</td>
<td>13.8</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>10</td>
<td>11</td>
<td>0.2169</td>
<td>0.3212</td>
<td>13.8</td>
</tr>
</tbody>
</table>

The Random Match Probability for this profile in the U.S. Caucasian population is **1 in 837 trillion (10^{12})**.
The Same 13 Locus STR Profile in Different Populations

1 in 837 trillion
1 in 0.84 quadrillion \((10^{15})\) in U.S. Caucasian population (NIST)
1 in 2.46 quadrillion \((10^{15})\) in U.S. Caucasian population (FBI)*
1 in 1.86 quadrillion \((10^{15})\) in Canadian Caucasian population*

1 in 16.6 quadrillion \((10^{15})\) in African American population (NIST)
1 in 17.6 quadrillion \((10^{15})\) in African American population (FBI)*

1 in 18.0 quadrillion \((10^{15})\) in U.S. Hispanic population (NIST)

These values are **for unrelated individuals** assuming no population substructure (using only \(p^2\) and 2 pq)


*http://www.csfs.ca/pplus/profiler.htm*
Approaches for “challenging” samples: perspectives for the future

- Limited sample material (highly degraded DNA)
  - mtDNA (in use for this purpose since mid-1990s due to high copy number per cell)

- Mixed male-female DNA
  - Y-chromosome STRs

- Degraded DNA
  - miniSTRs
  - SNPs (?)

Chapter 10 in *Forensic DNA Typing, 2nd Edition*
Different Inheritance Patterns

CODIS STR Loci

Autosomal
(passed on in part, from all ancestors)

Y-Chromosome
(passed on complete, but only by sons)

Mitochondrial
(passed on complete, but only by daughters)

Y-STRs can permit simplification of male DNA identification in sexual assault cases.

Female Victim DNA Profile

Male Perpetrator DNA Profile

DNA Profile from Crime Scene

Autosomal STR Profile

Y-Chromosome STR Profile

No signal observed


Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology
Y-Chromosome Haplotype Reference Database (YHRD)

http://www.yhrd.org

As of 12/17/04: **28,650 haplotypes**

**6,281 haplotypes**
with all US required loci

Commercial Y-STR kits exist to amplify all of the core loci in a single reaction (plus a few additional markers)

Run only with minimal haplotype

DYS19
DYS389I/II
DYS390
DYS391
DYS392
DYS393
DYS385 a/b

US haplotype requires 2 additional loci:
DYS438
DYS439
miniSTRs: new tool for degraded DNA

Smaller PCR products work better with low copy number or fragmented DNA templates

Conventional STR test (COfiler™ kit)

MiniSTR assay (using Butler et al. 2003 primers)
Report published in Nov 2000

Asked to estimate where DNA testing would be 2, 5, and 10 years into the future

Conclusions

STR typing is here to stay for a few years because of DNA databases that have grown to contain millions of profiles
Additional Resources


Content of STRBase Website

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

- ./str_fact.htm  STR Fact Sheets on Core Loci
- ./multiplx.htm  Multiplex STR Kit Information
- ./y_strs.htm   Y-Chromosome Information
- ./var_tab.htm  Variant Alleles Reported
- ./mutation.htm Mutation Rates for Common STRs
- ./str_ref.htm  Reference List with ~2,300 Papers
- ./training.htm Downloadable PowerPoints for Training
- ./validation.htm Validation Information
- ./miniSTR.htm  miniSTR Information
- ./address.htm  Addresses for Scientists
- ./NISTpub.htm  Publications & Presentations from NIST
Summary of Key Points

- STRs are highly variable genetic markers
- Core STR loci have been chosen to enable a common currency for use in national DNA databases
- STR kits permit co-amplification of up to 15 STRs plus amelogenin for sex-typing
- Capillary electrophoresis with fluorescence detection has become the method of choice for STR typing
- STR allele frequencies are used to estimate the rarity of a particular STR profile
- The core STR markers of today should remain in widespread use due to the millions of profiles already in DNA databases
- Y-STRs and miniSTRs will likely play a growing role in the future
Questions

Your turn, any questions?

>> Click on the Q&A tab, type your name & question, hit send and I will answer it live now!

John Butler
john.butler@nist.gov
301-975-4049
http://www.cstl.nist.gov/biotech/strbase

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