The Human Y-Chromosome: Background and Use in Forensic DNA Typing

John M. Butler
National Institute of Standards and Technology
CODIS Conference – October 23, 2006
Arlington, VA

Summary of 2006 CODIS Survey Questions Regarding Y-STRs

Questions #45a & #45b

- Is your lab using or validating Y-STRs?
  - 51 Yes (30%)
    - 28 Yfiler, 15 PowerPlex Y, some both kits
  - 114 No
  - 6 no response

Presentation Outline

- Background on human Y-chromosome
- Why Y is of interest in human identity testing
- Y-STR markers and kits available
- Different population databases and statistics for reporting matches

Different Inheritance Patterns

- 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)
Role of Y-STRs and mtDNA Compared to Autosomal STRs

- **Autosomal STRs provide a higher power of discrimination and are the preferred method whenever possible.**
- **Due to capabilities for male-specific amplification,** Y-chromosome STRs (Y-STRs) can be useful in extreme female-male mixtures (e.g., when differential extraction is not possible such as fingernail scrapings).
- **Due to high copy number,** mitochondrial DNA (mtDNA) may be the only source of surviving DNA in highly degraded specimens or low quantity samples such as hair shafts.

Value of Y-Chromosome Markers

J.M. Butler (2005) Forensic DNA Typing, 2nd Edition; Table 9.1

<table>
<thead>
<tr>
<th>Application</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forensic casework on sexual assault evidence</td>
<td>Male-specific amplification (can avoid differential extraction to separate sperm and epithelial cells)</td>
</tr>
<tr>
<td>Paternity testing</td>
<td>Male children can be tied to fathers in motherless paternity cases</td>
</tr>
<tr>
<td>Missing persons investigations</td>
<td>Patrilineal male relatives may be used for reference samples</td>
</tr>
<tr>
<td>Human migration and evolutionary studies</td>
<td>Lack of recombination enables comparison of male individuals separated by large periods of time</td>
</tr>
<tr>
<td>Historical and genealogical research</td>
<td>Surnames usually retained by males; can make links where paper trail is limited</td>
</tr>
</tbody>
</table>

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


Scenarios Where Y-STRs Can Aid Forensic Casework

- Sexual assaults by vasectomized or azoospermic males (no sperm left behind for differential extraction)
- Extending length of time after assault for recovery of perpetrator’s DNA profile (greater than 48 hours)
- Fingernail scrapings from sexual assault victims
- Male-male mixtures
- Other bodily fluid mixtures (blood-blood, skin-saliva)
- Gang rape situation to include or exclude potential contributors

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Disadvantages of the Y-Chromosome

- Loci are not independent of one another and therefore rare random match probabilities cannot be generated with the product rule; must use haplotypes (combination of alleles observed at all tested loci)

- Paternal lineages possess the same Y-STR haplotype (barring mutation) and thus fathers, sons, brothers, uncles, and paternal cousins cannot be distinguished from one another

- Not as informative as autosomal STR results
  - More like addition ($10 + 10 + 10 = 30$) than multiplication ($10 \times 10 \times 10 = 1,000$)

What has happened in the past few years...

- “Full” Y-chromosome sequence became available in June 2003; over 350 Y-STR loci identified (only ~20 in 2000)
- Selection of core Y-STR loci (SWGDAM Jan 2003)
- Commercial Y-STR kits released
  - PowerPlex Y (9/03), Yfiler (12/04)
- Many population studies performed and databases generated with thousands of Y-STR haplotypes
- Forensic casework demonstration of value of Y-STR testing along with court acceptance

Core Y-STR Characteristics

<table>
<thead>
<tr>
<th>STR Marker</th>
<th>Position (Mb)</th>
<th>Repeat Motif</th>
<th>Allele Range</th>
<th>Mutation Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>DY356</td>
<td>3.17</td>
<td>AGAT</td>
<td>8-17</td>
<td>0.05%</td>
</tr>
<tr>
<td>DY519</td>
<td>10.12</td>
<td>TAGA</td>
<td>10-19</td>
<td>0.20%</td>
</tr>
<tr>
<td>DY391</td>
<td>12.54</td>
<td>TCTA</td>
<td>6-14</td>
<td>0.40%</td>
</tr>
<tr>
<td>DY439</td>
<td>12.95</td>
<td>AGAT</td>
<td>8-15</td>
<td>0.38%</td>
</tr>
<tr>
<td>DY368 I/II</td>
<td>13.05</td>
<td>[TCTG] [TCTA]</td>
<td>8-17 / 24-34</td>
<td>0.20% / 0.31%</td>
</tr>
<tr>
<td>DY438</td>
<td>13.38</td>
<td>TTCTC</td>
<td>6-14</td>
<td>0.09%</td>
</tr>
<tr>
<td>DY390</td>
<td>15.71</td>
<td>[TCTA] [TCTG]</td>
<td>17-28</td>
<td>0.32%</td>
</tr>
<tr>
<td>DY385 a/b</td>
<td>19.19</td>
<td>GAAA</td>
<td>7-28</td>
<td>0.23%</td>
</tr>
<tr>
<td>DY392</td>
<td>20.97</td>
<td>TAT</td>
<td>6-20</td>
<td>0.05%</td>
</tr>
</tbody>
</table>

Positions in megabases (Mb) along the Y-chromosome were determined with NCBI build 35 (May 2004) using BLAT. Allele ranges represent the full range of alleles reported in the literature. Mutation rates summarized from YHRD (http://www.yhrd.org; accessed 6 Apr 2005).


http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Haplotype Databases for Y-STR Kits
http://www.promega.com/techserv/tools/pplexy/
http://www.appliedbiosystems.com/yfilerdatabase/

### PowerPlex Y

- 1311 Caucasians
- 325 Asians
- 894 Hispanics
- 1108 African Americans
- 366 Native Americans

**4,004 total**
(as of March 2005)

### Yfiler

- 1276 Caucasians
- 330 Asians
- 985 African Americans
- 106 Native Americans
- 105 Filipino
- 59 Sub-Saharan Africans
- 103 Vietnamese

**3,561 total**
(as of December 2004)

---

**Statistics with Y-STR Haplotypes**

Most labs will probably go with the **counting method** (number of times a haplotype is observed in a database) as is typically done with mtDNA results.

---

**Frequency Estimate Calculations**

In cases where a Y-STR profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

\[
p = \frac{X}{N}
\]

**Example:**

- 7 matches in 27,773

\[
p = \frac{7}{27,773} = 0.000252 \approx 0.00025\%
\]

An upper bound confidence interval can be placed on the profile’s frequency using:

\[
p \pm 1.96 \sqrt{\frac{(p)(1-p)}{N}}
\]

**Example:**

\[
p = 0.000252
\]

\[
1.96 \sqrt{\frac{(0.000252)(1-0.000252)}{27,773}} = 0.000187
\]

\[
0.000252 \pm 0.000187 = 0.000439 \approx 0.044\% (~1 \text{ in } 2270)
\]

---

**When there is no match...**

In cases where the profile has not been observed in a database, the upper bound on the confidence interval is

\[
1 - \alpha = \frac{1}{N}
\]

**Example:**

- 0 matches in 4,004

\[
1 - \alpha = 1 - \frac{0.05}{4,004} = 0.000748
\]

\[
0.000748 = 0.075\% (~1 \text{ in } 1340)
\]

If using database of 2,443, then the best you can do is 1 in 816.

---

**National U.S. Y-STR Population Database**

- Efforts underway at the **University of Central Florida** (with NIJ funding) to consolidate all U.S. data on Y-STR loci for population
- Data from ReliaGene, Promega, Applied Biosystems being gathered plus any forensic lab population sample data available

---

**The Meaning of a Y-Chromosome Match**

**Conservative statement for a match report:**

The Y-STR profile of the crime sample matches the Y-STR profile of the suspect (at xxx number of loci examined). Therefore, **we cannot exclude the suspect** as being the donor of the crime sample. In addition, we cannot exclude all patrilineal related male relatives and an unknown number of unrelated males as being the donor of the crime sample.

---

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

Funding from interagency agreement 2003-IJ-R-029 between the National Institute of Justice and the NIST Office of Law Enforcement Standards

NIST Human Identity Project Team – Leading the Way in Forensic DNA...

Tom Reid (DNA Diagnostics Center) – supplying the father-son samples for mutation rate analysis