Y-STR Data Interpretation

John M. Butler (NIST)
For the Virginia Department of Forensic Sciences
Science Advisory Committee

Richmond, VA
August 5, 2008

NIST Activities with Y-STRs

• SRM 2395 (Human Y Chromosome Standard)
• Characterized duplications and deletions
• Sequenced variant alleles
  – http://www.cstl.nist.gov/biotech/strbase/STRseq.htm
• Supplied ~20% of Yfiler 3561 database
• Measured mutation rates with Yfiler loci

26 publications since 2001 on NIST Y-chromosome work
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Value of Y-STRs to Forensic Casework

Y-STRs can extend range of potential solvable forensic cases

• Enabling detection of male DNA when mixed with excess female DNA
  – Sexual assaults by vasectomized or azoospermic males (no sperm left behind to enable differential extraction)
  – Fingernail scrapings from sexual assault victims
  – Other bodily fluid mixtures (blood-blood, skin-saliva)
  – Extending length of time after assault for recovery of perpetrator’s DNA profile (greater than 48 hours)

• Dealing with multiple male contributors
  – Gang rape situation to include or exclude potential contributors

• Gender clarification (with amelogenin Y null alleles)

• Extension of power of discrimination (with partial profiles)

Forensic Advantages of Y-STRs

• **Male-specific amplification** extends range of cases accessible to obtaining probative DNA results (e.g., fingernail scrapings, sexual assault without sperm)

• **Technical simplicity due to single allele profile**: can potentially recover results with lower levels of male perpetrator DNA because there is not a concern about heterozygote allele loss via stochastic PCR amplification; number of male contributors can be determined

• **Courts have already widely accepted STR typing**, instrumentation, and software for analysis (Y-STR markers just have different PCR primers)

• **Acceptance of statistical reports using the counting method** due to previous experience with mtDNA
Available Y-STR Loci, Kits and Databases

<table>
<thead>
<tr>
<th>Loci</th>
<th>Grouping (# Loci)</th>
<th>Available Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS19</td>
<td></td>
<td><a href="http://www.YHRD.org">http://www.YHRD.org</a></td>
</tr>
<tr>
<td>DYS389I</td>
<td></td>
<td>58,775 haplotypes (499 populations from around the world)</td>
</tr>
<tr>
<td>DYS389II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS392</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS385 a/b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS438</td>
<td>SWGDAM Core (11)</td>
<td><a href="http://www.usystrdatabase.org/13,906">http://www.usystrdatabase.org/13,906</a> haplotypes</td>
</tr>
<tr>
<td>DYS439</td>
<td></td>
<td><a href="http://www.YHRD.org">http://www.YHRD.org</a> 29,518 haplotypes</td>
</tr>
<tr>
<td>DYS437</td>
<td>PowerPlex Y (12)</td>
<td>Promega website: 4004 haplotypes</td>
</tr>
<tr>
<td>DYS448</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS456</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS458</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS635</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATA-H4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>~400 additional Y-STRs currently known</td>
<td>Hanson &amp; Ballantyne, Legal Med 2006;8(2):110-20</td>
</tr>
</tbody>
</table>

Hanson & Ballantyne, Legal Med 2006;8(2):110-20
Likely Success with Degraded DNA
7 Yfiler Loci <200 bp

AmpFSTR Yfiler Kit

6-FAM
DYS456  DYS389I  DYS390  DYS389II
VIC
DYS458  DYS319  DYS385a/b
NED
DYS393  DYS391  DYS439  DYS635  DYS392
PET
H4  DYS437  DYS438  DYS448

Loci unique to Yfiler kit

Yfiler Allelic Ladders

<200 bp
“Test 1” - Yfiler Profile

“Test2” - Yfiler Profile
Search Results with a Single Y-STR Locus

### Search Result with 1 locus (DYS456 allele 17)

<table>
<thead>
<tr>
<th>Ancestry</th>
<th># of Haplotypes</th>
<th>Number of Haplotypes (with Selected Alleles)</th>
<th>Frequency</th>
<th>Frequency Upper Bound (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>2138</td>
<td>152</td>
<td>0.073772</td>
<td>0.086089</td>
</tr>
<tr>
<td>Asian</td>
<td>596</td>
<td>38</td>
<td>0.089537</td>
<td>0.102384</td>
</tr>
<tr>
<td>Caucasian</td>
<td>2472</td>
<td>231</td>
<td>0.133900</td>
<td>0.147224</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1106</td>
<td>110</td>
<td>0.107905</td>
<td>0.125857</td>
</tr>
<tr>
<td>Native American</td>
<td>444</td>
<td>36</td>
<td>0.103004</td>
<td>0.131950</td>
</tr>
<tr>
<td>Total</td>
<td>6756</td>
<td>208</td>
<td>0.104900</td>
<td>0.111984</td>
</tr>
</tbody>
</table>

~1 in 9

### Search Result with 1 locus (DYS456 allele 16)

<table>
<thead>
<tr>
<th>Ancestry</th>
<th># of Haplotypes</th>
<th>Number of Haplotypes (with Selected Alleles)</th>
<th>Frequency</th>
<th>Frequency Upper Bound (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>2138</td>
<td>152</td>
<td>0.073653</td>
<td>0.077714</td>
</tr>
<tr>
<td>Asian</td>
<td>596</td>
<td>76</td>
<td>0.17832</td>
<td>0.208552</td>
</tr>
<tr>
<td>Caucasian</td>
<td>2472</td>
<td>205</td>
<td>0.318356</td>
<td>0.335729</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1106</td>
<td>322</td>
<td>0.292043</td>
<td>0.319841</td>
</tr>
<tr>
<td>Native American</td>
<td>444</td>
<td>92</td>
<td>0.218468</td>
<td>0.256903</td>
</tr>
<tr>
<td>Total</td>
<td>6756</td>
<td>1050</td>
<td>0.278198</td>
<td>0.286850</td>
</tr>
</tbody>
</table>

~1 in 3

Y-STR Haplotype Search Results

<table>
<thead>
<tr>
<th></th>
<th>YHRD Search</th>
<th>US YSTR Search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17 Yfiler</td>
<td>11 SWGDAM</td>
</tr>
<tr>
<td></td>
<td>9 MHL</td>
<td>7 Yfiler &lt;200bp</td>
</tr>
<tr>
<td>test1</td>
<td>0 of 3657</td>
<td>0 of 4163</td>
</tr>
<tr>
<td></td>
<td>17 of 29,518</td>
<td>5 of 3657</td>
</tr>
<tr>
<td></td>
<td>295 of 56,987</td>
<td>11 of 6601</td>
</tr>
<tr>
<td>test2</td>
<td>0 of 3657</td>
<td>0 of 4163</td>
</tr>
<tr>
<td></td>
<td>2 of 29,518</td>
<td>1 of 3657</td>
</tr>
<tr>
<td></td>
<td>3 of 56,987</td>
<td>6 of 6601</td>
</tr>
<tr>
<td>MCT</td>
<td>4 of 3657</td>
<td>9 of 4163</td>
</tr>
<tr>
<td></td>
<td>243 of 29,518</td>
<td>36 of 3657</td>
</tr>
<tr>
<td></td>
<td>1114 of 56,987</td>
<td>93 of 6601</td>
</tr>
</tbody>
</table>

~95% of worldwide Yfiler profiles (N=10,454) are unique
More loci are helpful in reducing “matches” (MHL – SWGDAM – Yfiler)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Sources of Yfiler Worldwide Population Data

28 published population studies with Yfiler data

- **ABI Database**
  - 3561 samples
  - N = 389 sons
  - N = 572 (w/laci)

- **Brazilian Study**
  - Pereira et al. (2007)
  - FSI 171:226-236
  - 500 males
  - 481 haplotypes (DC: 96%)
  - 466 unique
  - 5 geopolitical regions compared
  - \( \theta = 0.0013 \)

**6893 samples** (+3561 = 10,454)

- **6514 haplotypes** (discrimination capacity 94.5%)
- **6257 unique haplotypes** (96.0% unique)

Various Theta Values with a Partial Profile

\[
f = \theta + (1 - \theta)p = p + \theta(1 - p)
\]

For \( p < \theta \), \( \theta \) bounds the equation

<table>
<thead>
<tr>
<th>n</th>
<th>x</th>
<th>p</th>
<th>+95% CI</th>
<th>theta calc</th>
<th>theta</th>
</tr>
</thead>
<tbody>
<tr>
<td>6601</td>
<td>11</td>
<td>0.001666</td>
<td>0.00265</td>
<td>0.101500</td>
<td>0.1</td>
</tr>
<tr>
<td>6601</td>
<td>11</td>
<td>0.001666</td>
<td>0.00265</td>
<td>0.01650</td>
<td>0.01</td>
</tr>
<tr>
<td>6601</td>
<td>11</td>
<td>0.001666</td>
<td>0.00265</td>
<td>0.002964</td>
<td>0.0013</td>
</tr>
<tr>
<td>6601</td>
<td>11</td>
<td>0.001666</td>
<td>0.00265</td>
<td>0.001766</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Why a Y-STR theta correction is not needed (nor should be implemented)

• Full 17-locus profiles are unique ~95% of the time; partial profiles will have a higher degree of matches making a small theta irrelevant in many cases

• **No one else is doing it** (and no consensus in the community that it is required or necessary)
  – counting method is sufficiently conservative

• Would be too complicated to **accurately** employ on a routine basis – how would the “appropriate” population of interest be determined?

---

**Y-STR Expert Panels**
on Which I Have Served

• **ISFG DNA Commission**
  – Met in Nov 2004 in Berlin at Forensic Y Users Group
  – Drafted document in 2005 via email (nomenclature focus)
  – Published recommendations in 2006 in FSI & IJLM

• **SWGDAM Y-STR Subcommittee**
  – Initiated in July 2002; recommended loci in Jan 2003
  – Completed interpretation guidelines in Jan 2008
  – Guidelines to be published in FSC in Jan 2009
NIST Y-Chromosome Publications and Other Useful References
Handout for VA DFS SAB Meeting (August 5, 2008)


NIST Y-Chromosome Publications and Other Useful References
Handout for VA DFS SAB Meeting (August 5, 2008)

**ISFG Recommendations**


**STR Kits**


**Additional Loci**


**Databases**


Y-Chromosome Haplotype Reference Database (YHRD): [http://www.yhrd.org](http://www.yhrd.org)


**Population Variation and Data Interpretation**


**Mutation Rates**


**Duplication/Deletion**


**Joint Match Probability**
