Recent Developments in Y-Chromosome Analysis of Y-STRs and Y-SNPs

AAFS Y Chromosome Workshop
February 12, 2002
John M. Butler
National Institute of Standards and Technology

There is a growing interest in the Y-chromosome to aid forensic, paternity, and missing persons testing...

We need:
• Markers to type
• Population data
• Multiplex assays
• Standards

Outline of Presentation
• Standardizing Information on Y Chromosome Markers
  – STRBase and Other Internet Resources
• Development of New Y STR Multiplexes
  – Primer Design Strategies
  – NIST Y STR 20plex Results
• Y SNP Typing Technologies
  – Y SNP Multiplex Assays
  • NIST SRM 2395 with Y STR and Y SNP Data

History of Y STR Marker Discovery
• 1992 - DYS19 (Roewer et al.)
• 1994 - YCAI a/b, YCAII a/b, DXYS156 (Mathies et al.)
• 1996 - DY389I/II, DY390, DY391, DY392, DY393 (Roewer et al.)
• 1996 - DYF371, DY3425, DY3426 (Jobling et al.)
• 1997 - DY3288, DY3388 (Kayser et al.)
• 1998 - DY3385 a/b (Schneider et al.)
• 1999 - A7.1, A7.2, A10, C4, H4 (White et al.)
• 2000 - DY3434, DY3435, DY3436, DY3437, DY3438, DY3439 (Ayub et al.)
• 2000 - G09411, G10123 (de Knijff unpublished)
• 2001 - DY3441, DY3442 (Iida et al.)
• 2002 - DY3446, DY3447, DY3448, DY3449, DY3450, DY3452, DY3453, DY3454, DY3455, DY3456, DY3458, DY3459, DY3463, DY3464 (Redd et al.)

Blue labeled markers are in Y STR 20plex assay
Y STR Typing of Duplicated Regions

- A repeat
- B repeat
- Multiple primer binding sites occur giving rise to more than one PCR product for a given set of primers

A = B
A ≠ B

DY385 a/b and YCAII a/b
DYS389 I/II

Human Genome Data

- Human Genome Project has enlarged knowledge of Y chromosome...
- Public human genome sequence available at http://genome.ucsc.edu ("Golden Path")
- Entire Y chromosome sequence can be downloaded -- 51 Mbytes (last half is all "Ns" from heterochromatin region)
- ABCC STR sets -- total number of Y STRs that might be polymorphic
- We can position all Y STR markers on current Y chromosome sequence

Chromosomal Positions of Y STRs

Long Arm (q)
- DYS391
- DYS437
- DYS439
- DYS389 I/II
- DYS388
- DYS438
- DYS447
- DYS390
- H4
- DYS426
- YCAII a
- YCAII b
- DYS385 a
- DYS385 b
- A7.1
- DYS392
- DYS448

Short Arm (p)
- DYS393
- DYS19

Based on BLAT search from Aug 6, 2001 Human Genome Working Draft http://genome.ucsc.edu/

YCAII a and b: 885,555 bp apart
DYS385 a and b: 40,775 bp apart

Butler et al. 2002, submitted

Internet Resources for Y Chromosome Information

- http://www.ystr.org/usa/
- http://www.ystr.org/europe/
- http://www.ystr.org/asia/
- http://www.medfac.leidenuniv.nl/fldo/
- http://www.ystr.org/asia/
- http://www.ystr.org/asia/

Short Tandem Repeat DNA

Internet Database

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

Y-Chromosome STR Information Available

- Over 200 publications on Y STRs & SNPs cataloged
- Allele information on 16 Y STR loci
- Downloadable PowerPoint on Y STRs and Y SNPs
- Links to other Y-chromosome sites

Butler et al. 2002, submitted
Standard Way to Type “Extended Haplotype”

- GDB primers (original primers; not for multiplexing)
- 2-3 multiplexes for minimal haplotype
- DYS385 run separately sometimes
- YCAI run separately to obtain extended haplotype
- Different PCR conditions
- Primers often require titrations (10 fold difference in amounts) in order to establish reasonable balance between Y STR PCR products

Published primers were not designed to work together (originally used in singleplex PCRs)

First Effort at Designing a Y STR Multiplex Assay

Gusmão et al. 1999
(FORENSIC SCI INT, VOL. 106, PP. 163-172)

Higher level multiplexes are needed to improve the power of discrimination for Y chromosome DNA tests

Dye color
PCR product size range

Inefficient use of space across size range and dye color

Multiplex PCR Requires Combination of Many Primers

HPLC may be needed for improving purity following synthesis especially with dye labeled primers

- Quality control testing of oligonucleotides used in multiplex PCR reactions is essential
- Reverse-phase HPLC
- MALDI-TOF mass spectrometry
- Y STR 20plex contains 34 different primers

Tools for Multiplexing

- Primer design software
  - Visual Basic programs to check potential primer dimer formation
- Quality control testing of primers
- Rapid multiplex testing
Assay Development Goals

- Single tube amplification of all loci in minimal (9-loci) or extended (11-loci) European Y STR haplotypes
- Incorporation of additional polymorphic markers recently published
- No female amplification products (avoid X homology)
- Spacing between loci in same color to allow additional undiscovered alleles to be accommodated
- Similar concentration of primers to produce balanced amplification products
- Sensitivity to <500 pg male DNA with 28 cycle PCR (similar behavior to current STR kits)

The first demonstration of our Y STR 20plex was on July 5, 2001

Summary of Y STR Multiplex Assays

<table>
<thead>
<tr>
<th>Locus</th>
<th>Ref. Allele</th>
<th>NIST 20plex</th>
<th>NIST 10plex</th>
<th>Y-PLEX 6</th>
<th>Guernsey 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DYS390</td>
<td>15</td>
<td>157 bp</td>
<td>157 bp</td>
<td>157 bp</td>
</tr>
<tr>
<td>2</td>
<td>DYS389I</td>
<td>12</td>
<td>155 bp</td>
<td>155 bp</td>
<td>155 bp</td>
</tr>
<tr>
<td>3</td>
<td>DYS389II</td>
<td>29</td>
<td>275 bp</td>
<td>275 bp</td>
<td>275 bp</td>
</tr>
<tr>
<td>4</td>
<td>DYS390</td>
<td>24</td>
<td>217 bp</td>
<td>217 bp</td>
<td>217 bp</td>
</tr>
<tr>
<td>5</td>
<td>DYS391</td>
<td>11</td>
<td>169 bp</td>
<td>169 bp</td>
<td>169 bp</td>
</tr>
<tr>
<td>6</td>
<td>DYS392</td>
<td>12</td>
<td>260 bp</td>
<td>260 bp</td>
<td>260 bp</td>
</tr>
<tr>
<td>7</td>
<td>DYS393</td>
<td>12</td>
<td>121 bp</td>
<td>121 bp</td>
<td>121 bp</td>
</tr>
<tr>
<td>10</td>
<td>YCAII</td>
<td>23</td>
<td>150 bp</td>
<td>150 bp</td>
<td>150 bp</td>
</tr>
<tr>
<td>12</td>
<td>DYS388</td>
<td>12</td>
<td>157 bp</td>
<td>157 bp</td>
<td>157 bp</td>
</tr>
<tr>
<td>14</td>
<td>DYS437</td>
<td>16</td>
<td>164 bp</td>
<td>164 bp</td>
<td>164 bp</td>
</tr>
<tr>
<td>15</td>
<td>DYS438</td>
<td>10</td>
<td>226 bp</td>
<td>226 bp</td>
<td>226 bp</td>
</tr>
<tr>
<td>17</td>
<td>A7.1</td>
<td>10</td>
<td>113 bp</td>
<td>113 bp</td>
<td>113 bp</td>
</tr>
<tr>
<td>18</td>
<td>H4</td>
<td>12</td>
<td>136 bp</td>
<td>136 bp</td>
<td>136 bp</td>
</tr>
<tr>
<td>20</td>
<td>DYS395</td>
<td>9</td>
<td>145 bp</td>
<td>145 bp</td>
<td>145 bp</td>
</tr>
<tr>
<td>22</td>
<td>DYS396</td>
<td>12</td>
<td>157 bp</td>
<td>157 bp</td>
<td>157 bp</td>
</tr>
<tr>
<td>23</td>
<td>DYS397</td>
<td>12</td>
<td>159 bp</td>
<td>159 bp</td>
<td>159 bp</td>
</tr>
<tr>
<td>24</td>
<td>DYS438</td>
<td>12</td>
<td>164 bp</td>
<td>164 bp</td>
<td>164 bp</td>
</tr>
</tbody>
</table>

New Y STRs

- Discovered by Mike Hammer’s Group at the University of Arizona
- Will be included in future multiplex assays

Development Strategy for Y STR Multiplexes

- Careful definition of allele ranges
  - Literature searches of over 200 papers to locate all known alleles
  - Evaluation of diverse population samples to search for rare alleles (M. Hammer cell lines)
  - Permits markers to be packed together more closely in a single dye color
- Efforts to avoid polymorphic nucleotides in primer binding sites
  - Alignment of multiple GenBank entries
  - Permits primers to be designed with less risk of null alleles
- Incorporation of newly developed NIST multiplex assay design and testing tools

Need to Define Allele Range Well

An extreme “off-ladder” allele seen in YCC panel
Intelligent Primer Design

- Overall balance in calculated primer Tm
- Check for primer-primer interaction

Examples
- YCAII
  - improved primer Tm compared to commonly used set
- DYS391
  - efforts to avoid X homology
- DYS19
  - moving primers in an effort to improve efficiency resulted in discovery of duplicated region of the Y

Y STR 20plex Primers (Predicted T_m)

- Average: 59.1
- Std Dev.: 2.8

Screening for potential intramolecular hairpin and intermolecular primer-dimer formation

YCAII: Improving Primer T_m

- Tm = 43.0°C
- Tm = 59.2°C

DYS391: Avoiding the X Homolog

Female artifact problems seen with published and Y-PLEX 6 primers

Comparison of Various Primers on DYS19 and a Newly Discovered Duplicated Region of the Y Chromosome
PCR Products Seen Using Different Primer Sets for the DYS19 Locus

Schoske et al, in preparation

Schematic of Loci in NIST Y STR 20plex

Butler et al. (submitted)

Genotyper® 3.7 Macro (NT)

Butler et al. (submitted)

Comparison of Y STR 20plex to Commercial Y-Plex™ 6 Kit (ReliaGene)
Advantages of Y STR 20plex

- Single amplification of “extended haplotype”
  - European loci: 19, 385 I/II, 389I, 390, 391, 392, 393, YCAI a/b
  - Additional loci: 426, A7.1, H4, 388, 437, 438, 439, 447, 448
- Sensitive to <250 pg with 28 cycle PCR
- Male-specific with >100x female DNA
- 10 loci with amplicons less than 200 bp in size to aid results with degraded DNA

Y Chromosome Consortium (YCC) Cell Line Samples

74 males and 2 females from diverse world populations

- 11 !Kung
- 5 Pygmy
- 10 Bantu speakers
- 14 Europeans
- 5 Middle Easterns
- 3 Pakistanis
- 3 Chinese
- 3 Japanese

http://ycc.biosci.arizona.edu/

Source of YCC Samples

Problems with male lineages in population databases (YCC 6/7, 12/13, 15/16, 29/30, 49/50, 8/37) - really need detailed pedigree information

- 11 !Kung
- 5 Pygmy
- 10 Bantu speakers
- 14 Europeans
- 5 Middle Easterns
- 3 Pakistanis
- 3 Chinese
- 3 Japanese

- 74 male cell lines
- 2 females (YCC1 and 54)

Established in 1991 by Mike Hammer and Nathan Ellis

http://ycc.biosci.arizona.edu/

New Alleles for DYS385

We have the samples to quickly build useful allelic ladders
John M. Butler, AAFS 2002 Y Chromosome Workshop Presentation

**Tri-allelic Pattern seen at DYS385**

- Y-PLEX 6 kit

**Discovery of New DYS391 Variant Allele**

- Ladder allele 10 + 250.63
- YCC Panel Data with Y-PLEX 6 kit

**Plans with our Y STR Megaplex**

- Currently being tested in 2 other labs
  - Mecki Prinz - forensic samples with 4 dye 16plex
  - Mike Hammer - population studies with 5 dye 20plex
- Concordance studies with YCC sample panel
- Allele sequencing...developmental validation
- Determine Y STR haplotypes for Y SRM 2395
- First publication submitted with details on primers
- Construct Allelic Ladders/LSBs
- We are committed to seeing this assay commercialized into a kit so that it can be widely available

**Allelic Ladders vs. LSBs**

- Locus-Specific Brackets (LSBs) are artificial alleles that serve as internal sizing standards, one smaller and one larger in size than the common set of STR alleles
- The flanking region of the LSB is identical to the STR alleles giving them the same electrophoretic properties
- LSBs are used instead of allelic ladders to type samples
- Collaboration has been established with OligoTrail to produce LSBs for Y STR multiplexes (NIJ-funded project)

**Y SNPs**

- Single Nucleotide Polymorphisms and Other Biallelic Markers

**Y Bi-allelic Markers**

<table>
<thead>
<tr>
<th>Marker Name</th>
<th>Polymorphism</th>
<th>PCR Product Size (bp)</th>
<th>Restriction Enzyme</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAC (DYS287)</td>
<td>Alu insertion</td>
<td>135 YAP-Y (435 YAP+Y)</td>
<td>MspI</td>
<td>Hammer 1994</td>
</tr>
<tr>
<td>DYS-689</td>
<td>11-A</td>
<td>299</td>
<td>BsrBI</td>
<td>Westfall 1995</td>
</tr>
<tr>
<td>DYS-510</td>
<td>A-G</td>
<td>297</td>
<td>AluII</td>
<td>Kiviat 1996</td>
</tr>
<tr>
<td>SRY-287</td>
<td>C-T</td>
<td>291</td>
<td>BstNI</td>
<td>Yats 1997</td>
</tr>
<tr>
<td>SRY-287</td>
<td>A-G</td>
<td>299</td>
<td>HpaII</td>
<td>Sassani 1994</td>
</tr>
</tbody>
</table>

**Until 1997, only 11 Y biallelic markers were known...**

>200 Y SNPs discovered by Peter Underhill (Stanford) using DHPLC

3,723 Y SNPs listed


Y SNPs Discovered as Part of the Human Genome Project

245 Y SNPs typed
74 males (YCC cell lines)
153 haplogroups observed

This paper unifies previous haplogroup nomenclatures
Primer and other information for all 245 markers are included in supplementary material

Instrumentation Required for SNP Assays

SNaPshot
Luminex Beads

Multi-Color Capillary Electrophoresis (ABI 310 or 3100)
Luminex 100 Flow Cytometer

Primer Extension
Roche LightCycler

Approach to Y SNP Assays

• Use branch-defining Y SNPs
• Develop multiplex PCR with optimized markers
  – M96, M42, M45, M9, M89
  – Recent Genome Research paper spells out best markers against the YCC panel
• Develop 10-20plex Y SNP panel with SNaPshot and mass spec approaches
  – 9plex PCR and 10plex SNaPshot assay already demonstrated with coding region mtDNA markers
  – 10plex mass spec assay from control region mtDNA following amplification of entire D-loop

Multiplex PCR (5plex) for Y SNPs

Y SNP Results for 16 Standard Samples

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>M9 (C/G)</th>
<th>M42 (A/T)</th>
<th>M45 (A/T)</th>
<th>M89 (C/T)</th>
<th>M96 (G/C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male 1</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Male 2</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Male 3</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Male 4</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Male 5</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Male 6</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Male 7</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Male 8</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Male 9</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Male 10</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Male 11</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Male 12</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Male 13</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Male 14</td>
<td>G</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Male 15</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Female</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
John M. Butler, AAFS 2002 Y Chromosome Workshop Presentation
Enables rapid creation of virtual and/or physical allelic ladders

From Y STR 20plex data

SRM 2395 Candidate Sample Testing

NIST Y STR 20plex

Y SNP Results with Primer Extension and MALDI-TOF MS

Summary of Y Chromosome Work at NIST

- We are standardizing information on Y chromosome markers and making it available through STRBase
- We are developing new Y STR multiplex assays (Y STR 20plex and Y STR 10plex)
- We are evaluating SNP typing methodologies and developing Y SNP assays involving primer extension
- We are developing a Y Chromosome Standard Reference Material (SRM 2395)

Acknowledgments

Funding:
National Institute of Justice Grant #97-LB-VX-0003
Interagency Agreement between NIJ and NIST Office of Law Enforcement Standards

NIST Personnel:
John Butler (Project Leader)
Pete Vallone
Margaret Kline
Jan Redman
Rich Schoske (AU)
Gordon Spangler (AU)
Christian Rutberg (RPI)
Dave Duewer (Anal. Chem.)

Collaborators:
Mike Hammer and Alan Redd (U. Arizona) for Y STR sequences and samples
Mecki Prinz (NYC OCME)
Dave Carlson (Marligen) on Y SNP work with Luminex beads

Human Genome Project participants (Jim Kent-UCSC)