Presentation Outline

- Advantages of the Y-Chromosome
- Characteristics of Core Y-STR Loci
- Available Y-STR Kits
- Available Y-STR Databases
- Y-STR Population Studies
- Statistics with Y-STR Haplotypes
- New Y-STRs
- Locus Duplication and Deletion
- Y-SNP

What has happened in the past 5 years

- "Full" Y-chromosome sequence became available in June 2003; over 200 Y-STR loci identified (only ~20 in 2000)
- Selection of core Y-STR loci (SWGDAM Jan 2003)
- Multiple commercial Y-STR kits released
  - Y-PLEX 6,5,12 (2001-03), 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

“State of the Y STR Assay” in June 2000

From J.M. Butler talk June 1, 2000 at Chi "DNA Forensics" meeting (Springfield, VA)

- A number of multiplex reactions have been reported in the literature but Y STR multiplexes have not reached their potential...
- Very little PCR optimization to-date (most work has been done with the original PCR primer sequences)
- No commercial Y STR kit exists yet (therefore these markers remain inaccessible to the general forensic DNA community)
- New Y STR markers are becoming available which will greatly improve the power of discrimination between unrelated individuals (e.g., DYS385) and these will need to be incorporated into future multiplex sets

What is the state of the Y chromosome? An evolutionary marker comes of age

Mark A. Jobling & Chris Tyler-Smith
Nature Reviews Genetics (2003) 4, 598-612

- Until recently, the Y chromosome seemed to fulfill the role of juvenile delinquent among human chromosomes — rich in junk, poor in useful attributes, reluctant to socialize with its neighbors and with an inescapable tendency to degenerate. The availability of the near-complete chromosome sequence, plus many new polymorphisms, a highly resolved phylogeny and insights into its mutation processes, now provide new avenues for investigating human evolution. Y-chromosome research is growing up.

Abstract

- spitting
- incessant use of TV remote buttons
- if lost, cannot stop and ask for directions
- ability to recall facts about baseball/basketball/hockey/golf/etc.
- male pattern baldness
- congregates with other Y-chromosome bearers to do "guy things"
- Source of "Testosterone poisoning"

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
J.M. Butler, “State of the Y Chromosome” address

**Value of Y-Chromosome Markers**

<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>

**Advantages of the 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

**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

**Forensic Advantages of Y-STRs**

- 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

**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

**Utility of the Y-STR Typing Systems Y-PLEX™ 5 and Y-PLEX™ 6 in Forensic Casework and 11 Y-STR Haplogroup Database for Three Major Population Groups in the United States**

**Some Reported Casework Examples**

<table>
<thead>
<tr>
<th>Kit/Loci Used</th>
<th>Reference</th>
</tr>
</thead>
</table>

**Y-STRs in Casework**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>STR Profile</th>
<th>Other Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA obtained from a semen stain</td>
<td>Y-PLEX 6 profile</td>
<td>Profile matched suspect</td>
</tr>
<tr>
<td>DNA obtained from a blood sample</td>
<td>Y-PLEX 5 profile</td>
<td>Profile matched father</td>
</tr>
</tbody>
</table>

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Y-STRs can permit simplification of male DNA identification in sexual assault cases

<table>
<thead>
<tr>
<th>Profile Type</th>
<th>DNA Profile</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Victim DNA Profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male Perpetrator DNA Profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA Profile from Crime Scene</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


**PowerPlex Y Performance in Our Hands**

<table>
<thead>
<tr>
<th>Concentration</th>
<th>female DNA</th>
<th>Male:female DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ng</td>
<td>80x female</td>
<td>1:2</td>
</tr>
<tr>
<td>2 ng</td>
<td>500x female</td>
<td>1:250</td>
</tr>
<tr>
<td>1 ng</td>
<td>408x female</td>
<td>1:408</td>
</tr>
</tbody>
</table>


---

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


---

**“Full” Y-Chromosome Sequence**

Sequence analysis of the Y-chromosome has revealed that it is much more complex than previously thought.


---

**Y-Chromosome Sequence Published**


The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes.


---

**X-Chromosome Sequence Published**


The DNA sequence of the human X chromosome.


---

PCR primers for Y-STRs need to be carefully designed to avoid X-chromosome homology (e.g., DYS391 in Y-PLEX 6 kit).
DYS391 Primer Improvements

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

New PCR primers designed at NIST

Significant homology exists between X and Y sequences for the DYS391 locus

We have designed primers to anneal to regions that only appear on the Y chromosome (targeted X deletion regions)

These primers produce smaller PCR products and were adopted by Promega for their PowerPlex Y kit


Y STRs permit extension of possible reference samples in missing persons cases

Y-STR Typing of Duplicated Regions

“multi-copy loci”

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

Y-STR loci are often counted by the number of amplicons rather than the number of PCR primer pairs

Y-PLEX™ 6 results

History of Y STR Marker Discovery

1992 – (DYTS1) (Roewer et al.)
1994 - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)
1996 - DYTS10, DYTS16, DYTS31, DYTS32, DYTS33 (Roewer et al.)
1996 - DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 (Roewer et al.)
1996 - DYF371, DYS425, DYS426 (Jobling et al.)
1997 - DYS586, DYS388 (Kayser et al.)
1998 - DYS355 a/b, Schneider et al.)
1999 - A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4 (White et al.)
2000 - DYS434, DYS435, DYS436, DYS437, DYS439 (Ayub et al.)
2000 - G09411 (DYS462), G10123 (de Knijff unpublished)
2001 - DYF441, DYS442 (Iida et al.)
2002 - DYS443, DYS444, DYS445 (Iida et al.), DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS460 a/b, DYS463, DYS464 a/b/c (Rood et al.)
2002 – DYS468-DYS506 – 15 new Y STRs (Manfred Kayser, GDB entries)
2003 – DYS597-DYS645 – 10 new Y STRs (Manfred Kayser, GDB entries)


DYS385 a/b and YCAII a/b

Y-PLEX™ 6 results

Different Inheritance Patterns

CODIS STR Loci

Lineage Markers

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)


Important Y-STR Loci:

1992 - (DYTS1) (Roewer et al.)
1994 - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)
1996 - DYTS10, DYTS16, DYTS31, DYTS32, DYTS33 (Roewer et al.)
1996 - DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 (Roewer et al.)
1996 - DYF371, DYS425, DYS426 (Jobling et al.)
1997 - DYS586, DYS388 (Kayser et al.)
1998 - DYS355 a/b, Schneider et al.)
1999 - A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4 (White et al.)
2000 - DYS434, DYS435, DYS436, DYS437, DYS439 (Ayub et al.)
2000 - G09411 (DYS462), G10123 (de Knijff unpublished)
2001 - DYF441, DYS442 (Iida et al.)
2002 - DYS443, DYS444, DYS445 (Iida et al.), DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS460 a/b, DYS463, DYS464 a/b/c (Rood et al.)
2002 – DYS468-DYS506 – 15 new Y STRs (Manfred Kayser, GDB entries)
2003 – DYS597-DYS645 – 10 new Y STRs (Manfred Kayser, GDB entries)

1992 - (DYTS1) (Roewer et al.)
1994 - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)
1996 - DYTS10, DYTS16, DYTS31, DYTS32, DYTS33 (Roewer et al.)
1996 - DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 (Roewer et al.)
1996 - DYF371, DYS425, DYS426 (Jobling et al.)
1997 - DYS586, DYS388 (Kayser et al.)
1998 - DYS355 a/b, Schneider et al.)
1999 - A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4 (White et al.)
2000 - DYS434, DYS435, DYS436, DYS437, DYS439 (Ayub et al.)
2000 - G09411 (DYS462), G10123 (de Knijff unpublished)
2001 - DYF441, DYS442 (Iida et al.)
2002 - DYS443, DYS444, DYS445 (Iida et al.), DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS460 a/b, DYS463, DYS464 a/b/c (Rood et al.)
2002 – DYS468-DYS506 – 15 new Y STRs (Manfred Kayser, GDB entries)
2003 – DYS597-DYS645 – 10 new Y STRs (Manfred Kayser, GDB entries)


http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Selection of U.S. Core Loci:

- DYS19
- DYS385 a/b
- DYS389 I/II
- DYS390
- DYS391
- DYS392
- DYS393
- DYS438
- DYS439

SWGDAM Sub-Committee on the Y Chromosome

- Formed in July 2002
- Members:
  - Jack Ballantyne (UCF) – chair
  - Mecki Prinz (NYC) – co-chair
  - John Butler (NIST)
  - Ann Gross (MN)
  - Jill Smerick (FBI)
  - Sam Baechtel (FBI)
  - John Hartmann (Orange Co.)
  - Jonathan Newman (CFS)
  - Phil Kinsey (OR)
  - Gary Sims (CA DOJ)
  - Bruce Budowle (FBI) – removed in 2004

- U.S. CORE Y-STR LOCI selected in January 2003
- 60 sample set selected for screening markers and initial testing
- Testing of Y-PLEX 6 and Y-PLEX 5 kits in all labs
  - All results completed agreed with NIST results sent to participating labs in Dec 2002
- Jack Ballantyne’s lab and John Butler’s lab to examine additional Y-STR and Y-SNP markers

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>DYS393</td>
<td>3.17</td>
<td>AGAT</td>
<td>8-17</td>
<td>0.05%</td>
</tr>
<tr>
<td>DYS19</td>
<td>10.12</td>
<td>TACA</td>
<td>10-12</td>
<td>0.20%</td>
</tr>
<tr>
<td>DYS391</td>
<td>12.54</td>
<td>TCTA</td>
<td>14-16</td>
<td>0.40%</td>
</tr>
<tr>
<td>DYS392</td>
<td>12.95</td>
<td>AGAT</td>
<td>8-14</td>
<td>0.38%</td>
</tr>
<tr>
<td>DYS389 I/II</td>
<td>13.05</td>
<td>[TCTG] [TCTA]</td>
<td>17-24</td>
<td>0.20%</td>
</tr>
<tr>
<td>DYS438</td>
<td>13.38</td>
<td>TTTT</td>
<td>14-16</td>
<td>0.09%</td>
</tr>
<tr>
<td>DYS390</td>
<td>15.71</td>
<td>[TCTA] [TCTG]</td>
<td>18-28</td>
<td>0.32%</td>
</tr>
<tr>
<td>DYS385 a/b</td>
<td>19.19, 19.23</td>
<td>GAAA</td>
<td>7-28</td>
<td>0.23%</td>
</tr>
<tr>
<td>DYS392</td>
<td>20.97</td>
<td>TAT</td>
<td>6-20</td>
<td>0.05%</td>
</tr>
</tbody>
</table>

Example Y STR Fact Sheet from STRBase

Y-Chromosome Standard NIST SRM 2395

- Certified for all loci in commercial Y-STR kits
- SWGDAM recommended loci
- Y-PLEX 6
- Y-PLEX 5
- Y-PLEX 12
- PowerPlex Y
- Y-filer - adds DYS635 (C4), now sequenced

Helps meet FBI Standard 9.5 (and ISO 17025)...traceability to a national standard
Available Y-STR Kits

Y STR Allele Nomenclatures

Y STR Allele Nomenclatures

Commercial Y-STR Kits

Yfiler (Applied Biosystems)

Y-STR 20plex Amplification

Y-STR 20plex Amplification

Available Y-STR Kits

Y-STR 20plex Amplification

NIST 20plex Published Sept 2002

ALLELE SIZE RANGE AND LOCUS DYE COLORS

Y-STR 20plex Amplification

PowerPlex Y (Promega)

Commercial Y-STR Kits

Available Y-STR Kits
Validation Summary Sheet for PowerPlex Y

<table>
<thead>
<tr>
<th>Study Category</th>
<th>12plex Study</th>
<th>Description of Samples Tested (performed in 7 labs and Promega)</th>
<th>20plex Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Source</td>
<td>5 samples x 8 lates</td>
<td>45</td>
<td>10</td>
</tr>
<tr>
<td>Single Source (donor/con)</td>
<td>5 samples x 8 lates</td>
<td>45</td>
<td>10</td>
</tr>
<tr>
<td>Mixture Ratio</td>
<td>2 lates x 2 mixture series 1:1 alleles (2.5 ng male DNA, 2.5 ng female DNA)</td>
<td>102</td>
<td>205</td>
</tr>
<tr>
<td>Mixture Ratio (male/female)</td>
<td>2 lates x 2 mixture series 1:1 alleles (2.5 ng male DNA, 2.5 ng female DNA)</td>
<td>102</td>
<td>205</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>5 sets x 2 series x 5 samples</td>
<td>54</td>
<td>108</td>
</tr>
<tr>
<td>Nonspecifics</td>
<td>26 series</td>
<td>24</td>
<td>52</td>
</tr>
<tr>
<td>NIST NAT</td>
<td>6 components of NAT (250)</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Precision (ABI 3100 and ABI 377)</td>
<td>10 ladder replicates + 10 sample replicates + (8 ladders + 8 samples for 377)</td>
<td>38</td>
<td>76</td>
</tr>
<tr>
<td>Non-Protection Cases</td>
<td>85 cases with 153 samples</td>
<td>102</td>
<td>205</td>
</tr>
<tr>
<td>Stutter</td>
<td>412 samples used</td>
<td>412</td>
<td>824</td>
</tr>
<tr>
<td>Peak Height Ratio</td>
<td>N/A (except for DYS350 but no studies were noted)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Cycling Parameters</td>
<td>5 cycles (25/30/35/40/45) x 60-pulse series x 2 samples</td>
<td>60</td>
<td>120</td>
</tr>
<tr>
<td>Annealing Temperature</td>
<td>5 lates x 1 series (45/50/55/60/65°C) x 1 sample</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>Reduction volume</td>
<td>5 samples (5/5/10/15/20 μL) x (5 latters x 1 series)</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Thermal cycler test</td>
<td>4 replicates, per temperature x 1 sample</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Male-specificity</td>
<td>2 females x 1 ligation series (0.5 ng female DNA) x 5 samples each</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>TaqTag polymerase ligation</td>
<td>5 amounts (1, 1.5, 2, 2.5, 3 μL) x (5 latters x 1 series)</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>Primer pair ligation</td>
<td>5 amounts (0.1, 0.2, 0.3, 0.4, 0.5 μL) x (5 latters x 1 series)</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>Magnesium ligation</td>
<td>5 amounts (1, 1.5, 2, 2.5, 3 mg) x (5 latters x 1 series)</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

NIST Multiplexes for High-Throughput Y STR Typing

Laid the Groundwork for Applied Biosystems' Yfiler Y-STR Kit

11 U.S. core loci

<table>
<thead>
<tr>
<th>PCR Product Size (bp)</th>
<th>11plex</th>
</tr>
</thead>
<tbody>
<tr>
<td>450 bp</td>
<td>Yfiler Allelic Ladders</td>
</tr>
</tbody>
</table>

AmpF/STR® Yfiler™

Released by Applied Biosystems in Dec 2004

6-FAM

Blue

VIC

Green

NED

Yellow

PET

Red

LIZ

Orange

GS500-internal size standard

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Commercial Y-STR Kits Available

- ReliaGene Technologies (New Orleans, LA)
  - Y-PLEX™ 6: DYS19, DYS389I, DYS390, DYS391, DYS393, DYS385 a/b
  - Y-PLEX™ 5: DYS389II, DYS392, DYS438, DYS439
  - Y-PLEX™ 12: DYS19, DYS385 a/b, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

- Promega Corporation (Madison, WI)
  - PowerPlex® Y: DYS19, DYS385 a/b, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437

- Applied Biosystems (Foster City, CA)

- Serac (Bad Homburg, Germany)
  - genRES® DYSplex-1: DYS389II, DYS390, DYS391, DYS385 a/b, amelogenin
  - genRES® DYSplex-2: DYS19, DYS389II, DYS392, DYS393

- Biotype (Dresden, Germany)
  - Mentype® Argus Y-MH: DYS19, DYS385 a/b, DYS389II, DYS390, DYS391, DYS392, DYS393

Our Group at NIST Was Involved in Beta-Testing All U.S. Based Y-STR Kits

NIST SRM 2395 Component A

PowerPlex Y

Yfiler

Available Y-STR Haplotype Databases

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)

- US haplotype requires 2 additional loci: DYS438, DYS439

New Y-Chromosome Information Resources on STRBase

Yfiler Haplotype Database

http://www.appliedbiosystems.com/yfilerdatabase/

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
PowerPlex Y Haplotype Database
http://www.promega.com/techserv/tools/pplexy/

95 Caucasians 1311 Caucasians
284 Asians 325 Asians
630 Hispanics 894 Hispanics
577 African Americans 1108 African Americans
357 Native Americans 366 Native Americans

2,443 total 4,004 total

March 2005

Y-STR Population Studies

A few recent Y-STR population studies

<table>
<thead>
<tr>
<th>Population</th>
<th># Samples</th>
<th># Loci</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 North American groups</td>
<td>2,443</td>
<td>12</td>
<td>Budowle et al. (2005) FSI 150:1-15</td>
</tr>
<tr>
<td>91 European groups</td>
<td>12,700</td>
<td>7</td>
<td>Roezer et al. (2005) NAR 16:279-291</td>
</tr>
</tbody>
</table>

More than 200 Y-STR population studies have been published (most of this data is deposited in the YHRD – Y Chromosome Haplotype Reference Database)

PowerPlex Y Population Study

Twelve short tandem repeat loci: Y chromosome haplotypes. Genetic analysis on populations residing in North America


A compilation of Y-STR population data

Appendices in Rich Schoske’s Ph.D. dissertation; available on STRBase

- Source: over 200 published population data papers
- Helps define observed allele ranges, which aids in multiplex assay development (spacing between loci in the same dye color)
- Information is available to the community through the STRBase website – permits analysis of optimal markers for particular population
Richard Schoske Dissertation


- Worked at NIST from Nov 2000 to May 2003
- 270 page Ph.D. dissertation
- Entitled “The design, optimization and testing of Y chromosome short tandem repeat multiplexes.”
- Available for download on NIST STRBase website

Rich Schoske
PhD student from American University
Funded by Air Force
Graduated May 11, 2003

U.S. Population Data on 22 Y-STRs


Standard U.S. Population Dataset
http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm

260 Caucasians, 260 African Americans, 140 Hispanics, 3 Asians = 663 males

DNA extracted from whole blood (anonymous; self-identified ethnicities) received from Interstate Blood Bank (Memphis, TN) and Millennium Biotech Inc. (FL Lauderdale, FL)

To date: (~95,000 allele calls)

- Identifier (15 autosomal markers + Amelogenin) (10,608)
- Roche Linear Arrays (HV1/HV2 10 regions) (6,630)
- Y STRs 22 loci—27 amplicons (17,388)
- Y STRs 27 new loci (14,535)
- Yfiler kit 17 loci (11,237)
- Y SNPs 50 markers on sub-set of samples (11,498)
- Orchid 70 autosomal SNPs on sub-set (13,230)
- miniSTR testing new loci and CODIS concordance (9,228)
- mtDNA full-control region sequences by AFDIL

Statistical Calculations on Y-STR Data

- Locus (gene) Diversity = \( \frac{n(n-1)(1-\sum \pi_i^2)}{\sum n_i} \) where \( n \) is the number of samples in the dataset and \( \pi_i \) is the frequency of the \( i \)th allele
- Haplotype Diversity (HD) = \( \frac{n(n-1)(1-\sum \pi_i^2)}{\sum n_i} \) where \( n \) is the number of samples in the dataset and \( \pi_i \) is the frequency of the \( i \)th haplotype
- Random Match Probability (RMP) = 1 - HD
- Discrimination Capacity (DC) – total number of observed haplotypes divided by the total number of individuals in the dataset
- Unique Haplotypes (UH) – number of haplotypes that occur only once in the dataset

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Canadian Forensic DNA Technology Workshop
J.M. Butler, “State of the Y Chromosome” address
June 8, 2005

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.

Example Y-STR Haplotype

Core US Haplotype

- DYS19 – 14
- DYS389I – 13
- DYS389II – 29
- DYS390 – 24
- DYS391 – 11
- DYS392 – 14
- DYS393 – 13
- DYS385 a/b – 11,15
- DYS438 – 12
- DYS439 – 13

Matches by Databases

- YHRD (9 loci)
  - 7 matches in 27,773
- YHRD (11 loci)
  - 0 matches in 6,281
- ReliaGene (11 loci)
  - 0 matches in 3,404
- PowerPlex Y (12 loci)
  - 0 matches in 4,004
- Yfiler (17 loci)
  - 0 matches in 3,561

Y-Chromosome Haplotype Reference Database

www.YHRD.org

Release "15" from 2004-12-17 16:11:24

Minimal Haplotype Result

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 14
DYS393 – 13
DYS385 ab – 11,15

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:

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

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

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

\[ = 0.000252 \pm 0.000252 \]

\[ = 0.000252 \pm 0.000187 \]

\[ = 0.000439 \]

= 0.044% (~1 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

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

0 matches in 4,004

\[ 1-\alpha = 1-(0.05)\left(\frac{1}{4004}\right) = 0.000748 \]

= 0.075% (~1 in 1340)

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

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm 11
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 profiles 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.

Difficult Questions…

• Which database(s) should be used for Y-STR profile frequency estimate determination?

• Are any of the current forensic Y-STR databases truly adequate for reliable estimations of Y-STR haplotype frequencies?
  – Some individuals share identical Y-STR haplotypes due to recurrent mutations, not relatedness…
  – Is the database a random collection reflecting Y-STR haplotype frequencies of the population?
  – Is the Y-STR haplotype frequency relevant for the population of the suspect?

Issues raised by Peter de Knijff at his Promega meeting presentation (Oct 2004)

Conclusions from Peter de Knijff

From his presentation at the Promega meeting (Oct 2004)

A haplotype frequency taken from any Y-STR database should not be reported or seen as a random match probability

– Because all male relatives have the same haplotype

– Males can share haplotypes without being related

Database estimates are at most qualitative…

What Peter de Knijff Reports with a Y-STR Match

From his presentation at the Promega meeting (Oct 2004)

• The Y-STR profile of the stain matches with the suspect.

• Therefore, the suspect cannot be excluded as the donor of the stain.

• On the basis of this DNA evidence, I can also not exclude all paternally related male relatives of the suspect as possible donors of this stain.

• In addition, an unknown number of males from the same region cannot be excluded. A more accurate answer can only be obtained if (1) we have detailed knowledge of the population structure of the region of interest, (2) the Y-STR frequencies therein are known, and (3) we have knowledge about the family structure of the suspect.

New Y-STRs

More than 150 new Y-STR loci were characterized in June 2004. These new loci need to be studied in common sample sets including U.S. population groups in order to understand their ability to differentiate most common types and closely related individuals.

Most Common Type in Europeans

<table>
<thead>
<tr>
<th>Locus</th>
<th>Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS19</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>DYS389I</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>DYS389II</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>DYS390</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>DYS391</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>DYS392</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>DYS393</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>DYS385 a/b</td>
<td>11,14</td>
<td></td>
</tr>
</tbody>
</table>

1,116 matches in 28,650 samples without DYS385 (3.9%)

606 matches in 27,773 samples (2.1%)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Most Common Type seen in 22 NIST samples (3.7%) (from all 3 populations)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Minimal Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT84945</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
</tbody>
</table>

With SWGDAM US core loci (minimal haplotype=438,439)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Minimal Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT84945</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
</tbody>
</table>

With Promega’s loci (minimal haplotype=438,439,437)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Minimal Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT84945</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
</tbody>
</table>

With NIST 27 Y STRs (2 multiplexes) most common type breaks into 22 different groups (all samples differentiated)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Minimal Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT84945</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
<tr>
<td>MT94875</td>
<td>11,14-20-24-14-15-13-13-13</td>
</tr>
</tbody>
</table>

New Y-STR paper

June 2004 issue of American Journal of Human Genetics

A Comprehensive Survey of Human Y-Chromosomal Microsatellites

- Searched for all regions with ≥8 consecutive repeats and 2,3,4,5,6 bp repeat units
- Discovered 139 new polymorphic Y-STR loci (166 male-specific)
- Only studied so far in 8 different samples

With NIST Work with New Y-STR Loci

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Loci Not Pursued Further:
- Low Number of Alleles
- Primers Gave Artifacts in Female
  - DYS490 – duplicated and on chr X

Conversion to Allele Frequency Information

DYS463 Alleles Observed

Conversion to Allele Frequency Information

DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs

ISFG Guidelines for Y STRs
- Locus nomenclature should be DYS number if possible
- Allelic ladders should be used
- Allele nomenclature discussed...
ISFG Updated Y-STR Recommendations

Separating Brothers with 47 Y-STRs

- Two suspected brothers (ZT79338 and ZT79339) are part of our ~660 U.S. sample dataset at NIST.
- Thus far, we have generated 47 Y-STR allele calls on these samples.
- A mutation at DYS391 separates these individuals (one contains allele 11 and the other allele 10).
- These samples share autosomal STR alleles and contain identical mtDNA sequences.

Locus Duplication and Deletion

Events that impact Y-STR interpretation

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
**Different primers around DYS19 repeat result in selection of different regions of the Y-chromosome**

**Y-chromosome mapping**

<table>
<thead>
<tr>
<th>Locus</th>
<th># dup</th>
<th>&gt;1 repeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS19</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>DYS389I</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>DYS389II</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>DYS390</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>DYS391</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>DYS392</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DYS393</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>DYS385a/b</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>

*from www.yhrd.org, literature, and our work*

92% have single repeat difference

Since single-step mutations are most common, then single repeat spacing in duplicated alleles is expected

**Deciphering between a Mixture of Multiple Males and Locus Duplication**

- Note the number of loci containing >1 allele (other than multi-copy DYS385)
- Consider relative position on the Y-chromosome if multiple loci have two alleles
- See if repeat spread is >1 repeat unit
- Examine DYS385 for presence of >2 alleles

Locus duplication along the Y-chromosome is in many ways analogous to heteroplasmy in mitochondrial DNA, which depending on the circumstances can provide greater strength to a match between two DNA samples.

**CFS-1 with Yfiler**

DYS390 is deleted

**Sample PC0149 with Yfiler**

DYS392 is deleted
Deletions of some Y-STRs can be an inadvertent diagnosis of male infertility


- **AZFa deletion** (<1 in 100,000 men): expected to lack DYS389II, DYS437, DYS438, DYS439
- **AZFb deletion** (very rare): expected to lack DYS385 and DYS392
- **AZFc deletion** (1 in 4,000 men): expected to lack DYS464
- Possible that “Incomplete” haplotypes are not being submitted to the Y-STR haplotype databases
- Thus, Y-STRs are not neutral with respect to fertility information

Y-SNPs

*Recent Forensic Science Service Work with Y-SNPs*

Abstract

Marked differences in Y-SNP allele frequencies between continental populations can be used to predict the biogeographic origin of a man’s ancestral paternal lineage. Using 627 samples collected from individuals within the UK with pale-skinned Caucasian, dark-skinned Caucasian, African-Caribbean, South Asian, East Asian or Middle Eastern appearance we demonstrate that an individual’s Y-SNP haplogroup is also strongly correlated with their physical appearance. Furthermore, experimental evaluation of the Marligen SignetTM Y-SNP kit in conjunction with the Luminex 100 detection instrument indicates that reliable and reproducible haplogrouping results can be obtained from 1 ng or more of target template derived from a variety of forensic evidence types including, blood, saliva and post-coital vaginal swabs. The test proved highly male-specific with reliable results being generated in the presence of a 1000-fold excess of female DNA, and no anomalous results were observed during degradation studies despite a gradual loss of typable loci. Hence, Y-SNP haplogrouping has considerable potential forensic utility in predicting likely ethnic appearance.
Global Distribution of Y Haplogroups

Y-SNPs have been primarily typed in world populations

What haplogroups will be observed in U.S. populations?

Y-SNPs in U.S. populations

What haplogroups will be observed?

How specific will certain Y-SNPs be for a U.S. population group?

Forensic utility in comparison/addition to Y-STRs

Commercial kit (Marligen) 42 Y-SNPs

Medium sized multiplexes developed in-house (CE or MS)

Approaches to Y SNP Typing

Luminex 100 Flow Cytometer

Multi-Color Capillary Electrophoresis (ABI 310 or 3100)

Primer extension (SNAPshot assays)

Allele-specific hybridization (Marligen Signet Y SNP kit)

MT97125 (H3) in 94 C plate

MT97126 (A4) in 94 C plate

M172-G

M172-T

267 73.5

114 238.5

Allele fluorescent counts on Luminex system

SNP Detection by Hybridization

Luminex Bead Array Assay

Dye

PCR product

Identity of bead (probe)

~30 seconds to process each sample

Y SNP Assays Using Primer Extension (SNAPshot)

18 loci in 3 multiplex assays

Equal multiplexing done at both PCR and SNP levels (plexes)


Canadian Forensic DNA Technology Workshop
J.M. Butler, “State of the Y Chromosome” address

June 8, 2005

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Canadian Forensic DNA Technology Workshop
J.M. Butler, “State of the Y Chromosome” address

June 8, 2005

Y-SNPs Typed at NIST

- 42 SNPs + Amelogenin present in 5 multiplexes (commercially available kit from Marligen)
- 18 SNPs in 3 NIST-designed 6plexes (8 unique)
- 10 SNPs in 2 NIST-designed 5plexes (1 unique)
- 19 of the SNP sites overlapped…

Resulting in a total of 51 Y-SNPs

- 115 African Americans
- 114 Caucasians
- 95 Hispanics (presently typed for 10 Y-SNPs)

Potential Use for Y SNPs...

- Good ethnic separation (Caucasian)
- Good ethnic separation (African American)

Publication on U.S. Groups with Y-SNPs


Y-SNP Typing of U.S. African American and Caucasian Samples Using Allele-Specific Hybridization and Primer Extension

Different technologies yield the same Y-SNP type
- Full concordance was observed between hybridization and primer extension technologies on 18 different Y-SNPs (>3,800 allele calls)

Y-SNPs will have limited value for individualizing a sample
- 18 different types observed in 229 individuals

Current Y-SNPs appear to have limited value for ethnic differentiation in U.S. populations (with the exception of M2 that is only found in African Americans and not in Caucasians)

24 plates x 96 samples x 6plexes = >13,000 Y SNP allele calls

Observed Haplogroups in Two U.S. Populations

- 18 different haplogroups observed in 229 males

Y-SNP haplogroups for 115 African Americans
- Y-SNP haplogroups for 114 Caucasians

18 total Hgs; 5 shared

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

19
Y-STR Typing Conclusions

- Different technologies yield the same Y-STR type
- Full concordance was observed between hybridization and primer extension technologies on 18 different Y-STRs (+3,800 allele calls)
- Y-STRs will have limited value for individualizing a sample
- 18 different types observed in 229 individuals
- Current Y-STRs appear to have limited value for ethnic differentiation in U.S. populations
  - One exception: M2 only in African Americans; not in Caucasians

Summary of NIST Y-Chromosome Work

- Standardize information resources on Y-STRs and nomenclature for alleles
- Understand variation in U.S. populations so the best loci can be selected for commercial kits
- Construct multiplex assays to quickly evaluate loci
- Provide reference material for laboratory calibration (SRM 2395)

Y-Chromosome Publications from NIST (1)


Y-Chromosome Publications from NIST (2)


Y-Chromosome Publications from NIST (3)


Y-Chromosome Publications from NIST (4)


http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
International Forensic Y-User Workshops

- Next meeting (5th): Sept 26-30, 2006 (Innsbruck, Austria) – will also cover mtDNA
- 1st – Berlin, Germany June 1996
- 2nd – Berlin, Germany June 2000
- 3rd – Porto, Portugal Nov 2002
- 4th – Berlin, Germany Nov 2004

For more information, see: http://www.yhrd.org/index.html