Presentation Outline

- General NIST Projects
  - STRBase, etc.
- Y-Chromosome Work
  - SWGDAM Y-chromosome subcommittee
  - SRM 2395
  - New Y-STR loci under development
  - Y-SNPs
- Mitochondrial DNA Work
  - Coding region SNP assay development with AFDIL
  - LINEAR ARRAYS as an mtDNA screening tool
  - Standard Reference Material SRM 2392-I (Barbara Levin)
- Invitation to Participate in a New NIST Interlaboratory Study involving mixture interpretation

Current Areas of NIST Research Effort

- Standard Information Resources (STRBase information, training materials/review articles, validation standardization, calibration datasets)
- Interlaboratory Studies (Real-time PCR, mixture interpretation)
- Resources for “Challenging Samples” (miniSTRs for degraded DNA)
- Information on New Loci (Y-Chromosome, new STRs)
AutoDimer Primer Screening Program

Available for download from STRBase: http://www.cstl.nist.gov/biotech/strbase

A web-based interface is in development (similar to Primer3)

NIST Multiplexes for High-Throughput Y STR Typing

Standard U.S. Population Dataset

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

Y-Chromosome

SWGDAM, SRM 2395,
New Y-STR Loci, and Y-SNPs
Selection of U.S. Core Loci:
DYS19, DYS385 a/b, DYS389I/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)
  - John Hartmann (Orange Co.)
  - Sam Baechtel (FBI Lab)
  - Jill Smerick (FBI Lab)
  - Demris Lee (AFDIL)
  - Jonathan Newman (CFS-Toronto)
  - Phil Kinsey (OR)
  - Gary Sims (CA DOJ)
- 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 in the same sample set

European and U.S. Core Y-STR Loci

<table>
<thead>
<tr>
<th>Marker Name</th>
<th>Allele Range (repeat numbers)</th>
<th>Repeat Motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS19</td>
<td>10-19</td>
<td>TAGA</td>
</tr>
<tr>
<td>DYS385 ab</td>
<td>7-28</td>
<td>GAAA</td>
</tr>
<tr>
<td>DYS385 b</td>
<td>8-24</td>
<td>TCTG(TCTG)</td>
</tr>
<tr>
<td>DYS390</td>
<td>17-28</td>
<td>TCTA</td>
</tr>
<tr>
<td>DYS391</td>
<td>6-14</td>
<td>TCTG(TCTG)</td>
</tr>
<tr>
<td>DYS392</td>
<td>6-16</td>
<td>TAT</td>
</tr>
<tr>
<td>DYS393</td>
<td>8-17</td>
<td>AGAT</td>
</tr>
<tr>
<td>YCAII a/b</td>
<td>11-25</td>
<td>CA</td>
</tr>
<tr>
<td>DYS438</td>
<td>6-14</td>
<td>TTTTC</td>
</tr>
<tr>
<td>DYS439</td>
<td>8-15</td>
<td>AGAT</td>
</tr>
</tbody>
</table>

Y-Chromosome Standard NIST SRM 2395
- Human Y-Chromosome DNA Profiling Standard
- 5 male samples + 1 female sample (neg. control)
- 100 ng of each (50 µL at ~2 ng/µL)
- 22 Y STR markers sequenced
- 9 additional Y STR markers typed
- 42 Y SNPs typed with Marligen kit

Certified for all loci in commercial Y-STR kits:
- Y-PLEX 6
- Y-PLEX 5
- Y-PLEX 12
- PowerPlex Y

SWGDAM recommended loci:
DYS19, DYS385 ab, DYS389 ab, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

Y-filer - adds DYS635 (C4); now sequenced

We will continue to add information on new Y-STR loci as they are adopted by the community and put into commercial kits

Y-SNP Results on SRM 2395 from Marligen Signet™ Multiplexes (Luminex bead assay)
42 Y-SNPs measured across all samples

5 male components in SRM 2395 have 5 different Y-SNP backgrounds: R1b, J2, E3a, G, and I

http://www.fbi.gov
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Current NIST Y-STR Multiplex Assays

Extended Haplotype vs. US Haplotype

Extended
19, 388I, 390, 391, 392, 393, 385 a/b + YCAII a/b
US Haplotype: 19, 388I, 390, 391, 392, 393, 385 a/b + YCAII a/b

NIST Y-STR Loci Under Examination

Extended Haplotype vs. US Haplotype

Extended
19, 388I, 390, 391, 392, 393, 385 a/b + YCAII a/b
US Haplotype: 19, 388I, 390, 391, 392, 393, 385 a/b + YCAII a/b

Commercial Y-STR Kits

(YCAII a/b) DYS19 A7.1 (DYS460) DYS388/II A7.2 (DYS461) YCAII a/b DYS390 A10 DYS391 C4 DYS392 YCAII a/b DYS393 H4 DYS394 DYS395 DYS396 DYS397 DYS398 DYS399 a/b DYS400 DYS401 DYS402 DYS403 DYS404 DYS405 DYS406 DYS407 DYS408 DYS409 DYS410 DYS411 DYS412 DYS413 DYS414 DYS415 DYS416 DYS417 DYS418 DYS419 DYS420 DYS421 DYS422 DYS423 DYS424 DYS425 DYS426 DYS427 DYS428 DYS429 DYS430 DYS431 DYS432 DYS433 DYS434 DYS435 DYS436 DYS437 DYS438 DYS439 DYS440 DYS441 DYS442 DYS443 DYS444 DYS445 DYS446 DYS447 DYS448 DYS449 DYS450 DYS451 DYS452 DYS453 DYS454 DYS455 DYS456 DYS457 DYS458 DYS459 a/b DYS460 DYS461 DYS462 DYS463 DYS464 a/b/c/d DYS465 DYS466 DYS467 DYS468-DYS645 166 new Y STRs (Manfred Kayser GDB entries)

U.S. Population Data on 22 Y-STRs

Extended Haplotype vs. US Haplotype

Extended
19, 388I, 390, 391, 392, 393, 385 a/b + YCAII a/b
US Haplotype: 19, 388I, 390, 391, 392, 393, 385 a/b + YCAII a/b


Value of Additional Markers

Resolving the Minimal Haplotype Most Common Type

Unresolved with all 22 markers

647 males from 3 U.S. populations


http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
DYS464 is a powerful genetic marker but it will probably NOT be pursued in most human identity applications (it is being used extensively in genetic genealogy).

Y-STR Loci Under Examination

(Kayser et al.)

DYS485  DYS490  DYS495  DYS504  DYS506  DYS508  DYS520

(Kayser et al.)

DYS522  DYS525  DYS532  DYS533  DYS534  DYS540  DYS556

(Kayser et al.)

DYS557  DYS570  DYS575  DYS576  DYS594  DYS632  DYS643

These 22 loci have been examined in U.S. populations

New work (Manfred Kayser GDB entries)

Y-STR Loci Under Examination

(Kayser et al.)

DYS485  DYS490  DYS495  DYS504  DYS508  DYS520

(Kayser et al.)

DYS522  DYS525  DYS532  DYS533  DYS540  DYS556

(Kayser et al.)

DYS557  DYS570  DYS575  DYS576  DYS594  DYS632  DYS643

These 22 loci have been examined in U.S. populations

New work (Manfred Kayser GDB entries)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
First Screen with 95 Samples (Combo1 plate)

Low Number of Alleles

- DYS632-VIC
- DYS575-FAM

High Number of Alleles

- DYS449-NED
- DYS570-NED

Number of Alleles Seen with Various Y-STR Loci in Same Set of 95 U.S. Samples

<table>
<thead>
<tr>
<th>Set</th>
<th>Locus ID</th>
<th>Dye</th>
<th>Length Range (bp)</th>
<th># alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>DYS632</td>
<td>VIC</td>
<td>153-157</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>DYS575</td>
<td>FAM</td>
<td>221-225</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>DYS556</td>
<td>FAM</td>
<td>214-222</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>DYS641</td>
<td>NED</td>
<td>222-230</td>
<td>3</td>
</tr>
<tr>
<td>G</td>
<td>DYS540</td>
<td>NED</td>
<td>265-277</td>
<td>4</td>
</tr>
<tr>
<td>J</td>
<td>DYS522</td>
<td>VIC</td>
<td>359-370</td>
<td>4</td>
</tr>
<tr>
<td>A</td>
<td>DYS495</td>
<td>FAM</td>
<td>133-146</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>DYS594</td>
<td>VIC</td>
<td>264-289</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>DYS505</td>
<td>FAM</td>
<td>168-187</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>DYS508</td>
<td>VIC</td>
<td>180-200</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>DYS520</td>
<td>NED</td>
<td>181-201</td>
<td>6</td>
</tr>
<tr>
<td>E</td>
<td>DYS533</td>
<td>VIC</td>
<td>209-230</td>
<td>6</td>
</tr>
<tr>
<td>J</td>
<td>DYS446</td>
<td>FAM</td>
<td>298-325</td>
<td>6</td>
</tr>
<tr>
<td>H</td>
<td>DYS444</td>
<td>NED</td>
<td>299-328</td>
<td>7</td>
</tr>
<tr>
<td>B</td>
<td>DYS643</td>
<td>FAM</td>
<td>135-167</td>
<td>7</td>
</tr>
<tr>
<td>D</td>
<td>DYS534</td>
<td>NED</td>
<td>202-227</td>
<td>7</td>
</tr>
<tr>
<td>D</td>
<td>DYS557</td>
<td>VIC</td>
<td>196-219</td>
<td>7</td>
</tr>
<tr>
<td>H</td>
<td>DYS525</td>
<td>VIC</td>
<td>303-328</td>
<td>7</td>
</tr>
<tr>
<td>J</td>
<td>DYS532</td>
<td>NED</td>
<td>466-490</td>
<td>7</td>
</tr>
<tr>
<td>D</td>
<td>DYS576</td>
<td>FAM</td>
<td>178-205</td>
<td>8</td>
</tr>
<tr>
<td>A</td>
<td>DYS485</td>
<td>NED</td>
<td>138-164</td>
<td>8</td>
</tr>
<tr>
<td>H</td>
<td>DYS504</td>
<td>FAM</td>
<td>270-303</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>DYS463</td>
<td>FAM</td>
<td>227-277</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>DYS570</td>
<td>NED</td>
<td>251-288</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>DYS490</td>
<td>VIC</td>
<td>126-158</td>
<td>10</td>
</tr>
<tr>
<td>B</td>
<td>DYS449</td>
<td>NED</td>
<td>341-381</td>
<td>12</td>
</tr>
</tbody>
</table>

Loci Not Pursued Further...

- Low Number of Alleles
- DYS632
- DYS575
- Primers Gave Artifacts in Female
- DYS490 – duplicated and on chr X
- DYS504
- DYS522
- DYS557

14,535 types generated across 27 loci

Combo 1 plate
31 Caucasians
32 African Americans
32 Hispanics

Collecting data to define full allele size ranges and optimal loci prior to constructing new large multiplexes

DYS76
15 AAAG repeats

Comparison of Different Datasets to Evaluate Variability in Locus

Crude allelic ladders are being created simply by mixing genomic DNA samples possessing each of the observed alleles

We plan to construct new Y-STR megaplexes with the best loci (based on haplotype discrimination not gene diversity ranking)
Characterizing Variant Alleles for Y-STRs

A new section of STRBase will be created soon to catalog and characterize these variants.

DYS635 Variant Allele 21.3
(Y-GATA-C4)

Yfiler allelic ladder

[CTTA]8 C [CTTT]3

DYS643 Variant
Allele 11.1

NIST Precision Sizing Data from ABI 3100

[CTTT]3 C [CTTT]3

Missing T

DYS635 Variant Allele 21.3
(Y-GATA-C4)

Yfiler allelic ladder

[TCTA][TCTA][TCTA][TCTA][TCTA][TCTA][TCTA][TCTA]

C insertion

DYS392 Flanking Region Polymorphism

Normal sequence

C-to-G Mutation

PowerPlex Y and Yfiler kits are external to this polymorphism while Y-PLEX 5, which creates a smaller amplicon for DYS392, is internal and therefore not impacted...

DYS19 Flanking Region Duplication

Different primers around DYS19 repeat result in selection of different regions of the Y chromosome.

DYS19 Flanking Region Duplication

Entire region of Y-chromosome has likely been duplicated and then diverged

Most duplications have a single repeat spread in allele patterns

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
**Duplication and Divergence Model**

<table>
<thead>
<tr>
<th>Locus</th>
<th># dup</th>
<th>&gt;1 repeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS391</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>DYS392</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>DYS393</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>DYS394</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>DYS395a/b</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DYS395ab</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>

92% have single repeat difference

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

**Y-SNPs**

- U.S. population sample data collected on 229 individuals (115 African Americans + 114 Caucasians)
- Published in *J. Forensic Sci.* (July 2004)
- Mapped 50 SNPs on Y-chromosome
- Compared SNAPSHOT and Luminex assays
- Certified values for 42 loci are available with SRM 2395

**Forensic SNP Site now a part of STRBase**

**Y-SNP Information Cataloged on STRBase Website**

http://www.cstl.nist.gov/biotech/strbase/SNPs/YSNPs50.htm

**Publication on U.S. Groups with Y-SNPs**


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)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Observed Haplogroups in Two U.S. Populations

- African Americans
  - B*
  - E3a
  - R1b
  - G
  - E3b
  - E2
  - E1
  - B2a

- Caucasians
  - R1b6
  - R1a1
  - K*
  - N3
  - E3*
  - J2

18 total Hgs; 5 shared

Our Recent Y-Chromosome Work


New Interlaboratory Study

- "Paper challenge" (no lab work required)
- Purpose to determine "lay of the land" for current practices in solving STR profile mixtures and reporting the results
- Results will be discussed at ISFG (Sept 2005)
- Please pick up a handout pertaining to our SOLICITATION FOR PARTICIPATION...

Summary of NIST Y-Chromosome Work

- Development of new Y-STR multiplex assays (Y STR 20plex, 11plex, etc.)
- Evaluation of new Y-STR loci in U.S. populations
- Evaluation of SNP typing methodologies and development of Y-SNP assays involving primer extension and the SNaPshot kit
- Creation of a Human Y-Chromosome Standard Reference Material (SRM 2395)
- Standardization of information on Y-chromosome markers with internet accessibility (STRBase)

>45,000 Y-chromosome allele calls generated to aid studies on optimal markers for U.S. populations

NIST mtDNA Work

- Coding Region
  - mtSNP 11plex (minisequencing assay)
  - Developed with AFIL to resolve mtDNA most common types
  - Int. J. Legal Med., 2004; 118: 147-157

- Roche Linear Arrays
  - (probes for HVI/HVII)

- Automated washing/
  - Population Study

Interlaboratory Comparisons

- 60 data sets
- Comparing results from 8 different samples using 9 different methods

Concordance

0 = Quantifiler
1 = Alu RT-PCR
5 = BRCA1

Mixture Interpretation Study 2005 (MIX05)

- "Paper challenge" (no lab work required)
- Purpose to determine "lay of the land" for current practices in solving STR profile mixtures and reporting the results
- Results will be discussed at ISFG (Sept 2005)
- Please pick up a handout pertaining to our SOLICITATION FOR PARTICIPATION...

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

NIST Project Team:
John Butler
Pete Vallone
Margaret Kline
Jan Redman
Amy Decker
Mike Coble
Richard Schoske (former member)

Funding:
Interagency Agreement between National Institute of Justice and NIST Office of Law Enforcement Standards

Collaborators:
Tom Parsons (AFDIL)
Jodi Irwin (AFDIL)
Sandy Calloway (Roche)

This presentation available as pdf file from
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm