New Autosomal and Y-Chromosome STR Loci: Characterization and Potential Uses

John M. Butler, Carolyn R. “Becky” Hill, Amy E. Decker, Margaret C. Kline, Peter M. Vallone
National Institute of Standards and Technology

18th International Symposium on Human Identification
Hollywood, CA
October 4, 2007

Questions to Be Addressed

• Why consider new STR loci?
• What has NIST accomplished with new autosomal STRs?
• Is there value in examining additional Y-STRs?
• Where can one learn more about these topics?

Aren’t the Current STR Loci Good Enough?

• Depends on the question being asked...
• For general forensic matching of evidence to suspect, the 13 CODIS STR loci are sufficient
• For other human identity/relationship testing questions, more autosomal or Y-STR loci can be beneficial or even necessary

How Would Additional STR Loci Be Useful?

• Databases: More loci to help resolve relatives in growing national DNA databases (UK went from 6 to 10 STRs in 1999; future Pan-European database will include >10 loci)
• Casework: Obtaining additional information with degraded DNA samples (miniSTRs); rapid screening of multiple crime scene samples
• Identity/Relationship Testing: Kinship analysis, parentage testing, complex criminal paternity, missing persons/mass disasters, immigration testing

Why consider new STR loci?

Aren’t the current core loci good enough?

How would additional STRs be useful?

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Call for More Loci in Situations Involving Relatives

- **Missing Persons** and Disaster Victim Identification (kinship analysis)
- Immigration Testing (often limited references)
  - Recommendations for 25 STR loci
- Deficient Parentage Testing
  - often needed if only one parent and child are tested

Relationship testing labs are being pushed to answer more difficult genetic questions…and we want to make sure the right tools are in place.

How are genetic loci introduced and adopted by the forensic/HID community?

Steps in Adopting Genetic Markers

- Loci Described
  - Assay Construction
  - Kit Development
- Population Study
  - Kit Testing
  - Release to Community
- Information Gathered
  - Internal Validation
  - Use in Casework
  - Court Presentation/Acceptance
- STRBase website
  - Research
  - Development
  - Forensic Application
  - Forensic Labs

History Plays a Role…

**J. Forensic Sci. 2006; 51(2): 253-265**

Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing

- Only 17 STRs, which were available from Applied Biosystems and Promega in kit or prototype kit form (in 1996-97), were evaluated as part of the selection process for the 13 CODIS core loci
- Human Genome Project has increased knowledge…now thousands of STRs are known

Steps in Adopting Genetic Markers

Role of the NIST Human Identity Project Team

Justice for All Act of 2004

- If additional loci are desired as core or supplementary loci on the national DNA database, the FBI must inform Congress six months prior to doing so...
  - "REPORT TO CONGRESS- If the Department of Justice plans to modify or supplement the core genetic markers needed for compatibility with the CODIS system, it shall notify the Judiciary Committee of the Senate and the Judiciary Committee of the House of Representatives in writing not later than 180 days before any change is made and explain the reasons for such change." (Section 203f)

What are important characteristics to consider in new loci?
Primary Characteristics in New STRs

- Genomic position
  - Adequate spacing from other (and current) loci to enable product rule use with autosomal markers
- Avoid known disease genes or linkage
  - To protect privacy concerns
- Polymorphic content (high heterozygosity)
  - More variable markers mean less can be used to reach desired rarity in full profile

Valuable Characteristics in New STRs

- Span/Range of observed alleles
  - Impacts electrophoretic real-estate
  - Tighter range makes differential amplification less likely
- Clean flanking region
  - To enable primer design near repeat (miniSTRs)
- Mutation rate known when trying to address multi-generational questions
- Provides benefit to haplotype resolution (Y-STRs)

Steps We Use in Characterizing New Loci

- Select genetic loci
- Design primers – optimize multiplex assay
- Type population samples to examine variation
- Sequence alleles to establish nomenclature
- Develop bins and panels for genotyping
- Construct allelic ladders
- Evaluate RMP or ability to separate common types
- Perform mutation rate studies
- Perform concordance studies (when applicable)
- Calibrate genotypes with NIST SRM components
- Work with companies/collaborators
- Publish details on loci and assays

Characterization of New Autosomal Loci

(miniSTR D12ATA63)

<table>
<thead>
<tr>
<th>Allele</th>
<th>Caucasian (N = 260)</th>
<th>African American (N = 259)</th>
<th>Hispanic (N = 140)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.0019</td>
<td>0.0154</td>
<td>0.0026</td>
</tr>
<tr>
<td>10</td>
<td>0.135</td>
<td>0.1523</td>
<td>0.0026</td>
</tr>
<tr>
<td>11</td>
<td>0.2154</td>
<td>0.1504</td>
<td>0.0268</td>
</tr>
<tr>
<td>12</td>
<td>0.0017</td>
<td>0.0071</td>
<td>0.0096</td>
</tr>
<tr>
<td>13</td>
<td>0.0679</td>
<td>0.0521</td>
<td>0.0981</td>
</tr>
<tr>
<td>14</td>
<td>0.2643</td>
<td>0.3340</td>
<td>0.1615</td>
</tr>
<tr>
<td>15</td>
<td>0.1004</td>
<td>0.2214</td>
<td>0.1786</td>
</tr>
<tr>
<td>16</td>
<td>0.1500</td>
<td>0.1525</td>
<td>0.1004</td>
</tr>
<tr>
<td>17</td>
<td>0.0036</td>
<td>0.0154</td>
<td>0.0173</td>
</tr>
<tr>
<td>18</td>
<td>0.0173</td>
<td>0.1564</td>
<td>0.0173</td>
</tr>
<tr>
<td>19</td>
<td>0.1385</td>
<td>0.1004</td>
<td>0.2154</td>
</tr>
</tbody>
</table>

Heterozygosity Values

- U.S. Caucasian: 0.842
- African American: 0.788
- U.S. Hispanic: 0.879

Unused Chromosomal Locations

(absolute to CODIS 13 STRs)

Plenty of room for additional loci that would be unlinked from current core STR loci

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
To Appear in Jan 2008 Issue of J. Forensic Sci.

- Primer sequences, GeneMapper bins and panels, genotypes on common samples, and allele frequency information already available on STRBase

How much DNA is required to obtain results with these new loci?

Assay Performance

- Our multiplex assays are designed to perform similarly to commercial kits
  - PCR Reaction (buffer, fluorescent dyes, volume)
  - PCR thermal cycling conditions
  - Work robustly on 0.5 to 1 ng of template DNA (or lower)
- Multiple miniplexes and a single megaplex developed to study 26 autosomal STRs

Multiple Miniplexes

- 26 characterized loci divided into nine 3plexes
- One locus per dye color
- Allelic ladders created
- Amplicons <140 bp
- miniSTRs
- Work with 100 pg DNA
  - For degraded samples (bones in missing persons cases)
    - NC = Non-CODIS or non-core

Single Megaplex

- So far 22 STRs and amelogenin in single multiplex (Eventual goal to have all 26 loci)
- Multiple loci in four dye channels
- Amplicons 70 to 400 bp
  - (No longer ‘miniSTRs’)
  - Typically use 1 ng DNA
- For reference samples (a missing person’s relatives)
  - “Autoplex” or “miniMegaplex”
  - All loci unlinked from core (CODIS) STRs

How does this megaplex perform?
Evaluation of Autoplex (23plex)

- **660 U.S. population samples**
  - U.S. Caucasian, African American, Hispanic
  - Concordance testing compared to miniSTR results

- **790 father/son samples**
  - U.S. Caucasian, African American, Hispanic, Asian
  - Mutation rate determination

- **12 samples for extended family testing**

> **1450 samples examined so far**

Concordance Study to Check for Null Alleles

Use of non-overlapping primers permits detection of allele dropout

“Autoplex” vs miniSTRs

- 639 samples compared
- Total types (639 x 22 loci): 14,058
- 28 types discordant (0.20%)*
- **99.80% concordance**

*Conclusions: (1) Our PCR primers have been well-designed and have almost no primer binding site mutations. (2) Roughly half of dropout is from megaplex primers – flanking regions near STR repeat do not appear to have a higher level of mutation

Mutation Rates Measured for New STRs

- **395 father/son pairs**
  - 22 STR loci examined
  - 8690 allelic transfers
  - **Only 6 mutations** were observed in total
  - 0.069%
  - (2-3 times less than typical 0.2% for common STRs)

**Conclusions**: Mutation rates are lower than commonly used STRs likely due to selection of loci for miniSTR application with tighter allele ranges, more moderate heterozygosities, and more stable flanking regions.

Population Data on New STRs

- **~660 samples** with three major U.S. populations on all 26 autosomal STR loci
  - Available on STRBase
      Allele_Frequencies_for_26miniSTRs.pdf

- **>3,000 samples tested world-wide** (Spain, Italy, Japan, Malaysia, Korea) on **first 6 loci** (NC01 & NC02)
  - D2, D10, D22 now recommended European loci

**Can these new STRs help in missing persons cases or other forms of relationship testing?**
**Extended Family Sample Testing**

- **Grandparents/children**
- **Aunt/Niece**
- **Mother/Child with mutation**
- **Cousins**
- **Siblings**

**Comparison of Likelihood Ratios**

<table>
<thead>
<tr>
<th>Relationship Examined</th>
<th>15 STRs (Identifiler, ID15)</th>
<th>ID15 + Autoplex 22 STRs = 37 loci (A37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother/Child* (with single mutation)</td>
<td>0.214</td>
<td><strong>5,200,000</strong> Extra loci help...</td>
</tr>
<tr>
<td>Siblings</td>
<td>477</td>
<td><strong>113,000</strong> Extra loci help...</td>
</tr>
<tr>
<td>Uncle/Nephew</td>
<td>824</td>
<td><strong>247,000</strong> Extra loci help...</td>
</tr>
<tr>
<td>Cousins</td>
<td>0.45</td>
<td><strong>2.25</strong></td>
</tr>
<tr>
<td>Grandparents/Grandchildren</td>
<td>0.53</td>
<td><strong>1.42</strong></td>
</tr>
</tbody>
</table>

**Y-CHROMOSOME STRs**

Are Y-STRs more sensitive than autosomal STRs?

**Are Y-STRs More Sensitive?**

- Y-chromosome markers (kits) are more selective as they offer male-specific amplification but the loci (kits) themselves are NOT more sensitive.

- Y-STRs have the same stochastic limitations with low-level DNA as autosomal markers

- However, allele dropout of heterozygote sister alleles (false homozygosity) is not an issue with single-copy Y-STRs

**What Y-STR loci and kits are commonly used today?**

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
J.M. Butler – Promega 2007 Talk
New Autosomal and Y-Chromosome STR Loci

Review of Available Y-STR Loci and Data

<table>
<thead>
<tr>
<th>Locus</th>
<th>Grouping (# Loci)</th>
<th>Available Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS51</td>
<td></td>
<td><a href="http://www.YHRD.org">http://www.YHRD.org</a></td>
</tr>
<tr>
<td>DYS389I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS390</td>
<td></td>
<td>50,867 haplotypes</td>
</tr>
<tr>
<td>DYS388</td>
<td></td>
<td>(464 populations from around the world)</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>DYS395 a/b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS425</td>
<td></td>
<td>SWGDAM Core (11)</td>
</tr>
<tr>
<td>DYS437</td>
<td></td>
<td>PowerPlex Y (12)</td>
</tr>
<tr>
<td>DYS446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS456</td>
<td></td>
<td>Yfiler (17)</td>
</tr>
<tr>
<td>DYS458</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYS459</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATA-H4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

~400 additional Y-STRs currently known
Hanson & Ballantyne, Legal Med 2006;8(2):110-20

What is NIST doing with Y-STRs?

NIST Activities with Y-STRs
- SRM 2395 (Human Y Chromosome Standard)
- Characterizing duplications and deletions
- Sequencing variant alleles
  - http://www.cstl.nist.gov/biotech/strbase/STRseq.htm
- Supplied ~20% of Yfiler 3561 database
- Concordance studies between Yfiler and NIST 20plex and 11plex assays
  - 22 publications since 2002 on NIST Y-chromosome work

NIST Work with Additional Y-STR Loci
- Studies of Locus Variation
  - 37 Y-STRs have been examined in all 665 NIST U.S. population samples and 92 Y-STRs in a subset (32 C, 32 AA, 31 H) using previously published primers and 3-5plexes
- Analysis of Mutation Rates
  - 389 father/son pairs with 17 Yfiler loci
- Further characterization of SRM 2395 components
  - To enable calibration with additional Y-STRs
- Defining allele nomenclature on 144 Y-STRs
  - To aid on-going genetic genealogy work

Publication with Additional Y-STR Loci


Announcement of population data
Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples
John M. Butler*, Amy E. Decker, Peter M. Valborne, Margaret C. Klime
Biocenter, Division of Medical and Biological Science, Oslo University, Norway
Received 2 January 2005, accepted 10 February 2005


Are there advantages to typing additional loci beyond the PowerPlex Y 12 or the Yfiler 17 Y-STRs?

Full 37 locus haplotypes available on STRBase:

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Lessons Learned from NIST Data Set

- Some Y-STRs that are more useful than others in sub-dividing common haplotypes (e.g., DYS576)
- You don’t gain much by typing additional Y-STRs (most unresolved types only occur twice)
- 95% of 17 locus Yfiler haplotypes are unique

Sources of Yfiler Worldwide Population Data
28 published population studies with Yfiler data

- 6893 samples
- 6514 haplotypes (discrimination capacity 94.5%)
- 6257 unique haplotypes (96.0% unique)

What additional population data exists with Yfiler?
And how does it compare to our NIST data?

Subdividing Unresolved Yfiler Haplotypes (1)

Lessons Learned from NIST Data Set

- Some Y-STRs that are more useful than others in sub-dividing common haplotypes (e.g., DYS576)
- You don’t gain much by typing additional Y-STRs (most unresolved types only occur twice)
- 95% of 17 locus Yfiler haplotypes are unique

Sources of Yfiler Worldwide Population Data
28 published population studies with Yfiler data

- 6893 samples
- 6514 haplotypes (discrimination capacity 94.5%)
- 6257 unique haplotypes (96.0% unique)

What additional population data exists with Yfiler?
And how does it compare to our NIST data?

Subdividing Unresolved Yfiler Haplotypes (1)
What do mutation rates look like for Y-STR markers?

Y-STR Mutation Rates Measured at NIST

- **389 father/son sample pairs**
  - U.S. Caucasians, African Americans, Hispanics and Asians
- **17 Y-STR loci** in the Yfiler kit
- **24 differences** between father and son
  - 13 mutations resulted in the gain of a repeat in the son
  - 11 resulted in a loss of a repeat
- **All single step repeat mutations**
  - except a two repeat loss at Y-GATA-H4
- **2 sample pairs were found to have two mutations**
  - African American pair: mutations at DYS458 and DYS635
  - Asian pair: mutations at DYS439 and Y-GATA-H4
- Also observed 4 duplications, 1 triplication, and 4 deletions that were seen in both father and son


What are the “best” additional Y-STRs?

The Next Best Y-STRs...Beyond the Kits

- **Decker et al. (2007) FSI Genetics 1:215-217**
  - DYS449, DYS505, DYS522, DYS532, DYS534, DYS570, DYS576
- **Hanson and Ballantyne (2007) PLoS ONE 8:e688**
  - DYS444, DYS446, DYS449, DYS459ab, DYS481, DYS508, DYS512, DYS527ab, DYS549, DYS562, DYS570, DYS576, DYS607, DYS627
  - DYS447, DYS449, DYS481, DYS570, DYS576

These loci are useful for subdividing common types and lineage testing...

To Summarize...

**Autosomal STRs**
- **26 unlinked loci** have been characterized and we have developed multiple miniplexes and a megaplex (23plex)
- Additional loci show value with relationship testing
- **NIST SRM 2391b** will include information on additional autosomal STR loci

**Y-Chromosome STRs**
- Studies at NIST and worldwide show ~95% of observed 17 locus Yfiler profiles are unique
- Additional loci can help with common types
- **NIST SRM 2395** will include information on additional Y-STR loci

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

Thank you for your attention...
Our team publications and presentations are available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

john.butler@nist.gov
301-975-4049

Collaborators
Mike Coble (now AFDIL)
- early miniSTR work
Tom Reid (DDC)
- father/son samples