STRs vs. SNPs: Thoughts on the Future of Forensic DNA Testing

John M. Butler
Michael D. Coble and Peter M. Vallone
U.S. National Institute of Standards and Technology
International Symposium on the Forensic Sciences (ANZFSS 2006)
April 4, 2006 – Fremantle, Australia

Presentation Outline

- Why consider SNPs for human identity testing?
- Work with SNPs at NIST
- Recent work with SNPs by others
- Direct comparisons of SNPs and STRs
- miniSTR work at NIST
- Score card: SNPs vs. STRs/miniSTRs

Reasons Often Given for Considering SNPs in Human Identity Testing…

- Use on degraded samples (WTC), low copy number, or telogenic (shed) hairs
- Lower mutation rate (Paternity testing)
- Easier data interpretation (no microvariants or stutter products)
- Amenable to high throughput analysis

Issues to be Addressed with SNPs

- Power of Discrimination
  - How many SNPs = 1 STR?
- Multiplex-ability (robust 50plex < 1ng DNA?)
- Population databases
- Many different platforms for SNP typing
- Unique interpretation issues – mixtures
- Validation
- Sensitivity
- Assay cost

Possible Allele Combinations

<table>
<thead>
<tr>
<th>STRs</th>
<th>SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGCTA(GATA)GACTAC</td>
<td>ATGCTA(C/T)GACTAC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>7.7, 7.8, 8.8</td>
</tr>
<tr>
<td>9</td>
<td>7.9, 8.9, 9.9</td>
</tr>
<tr>
<td>10</td>
<td>7.10, 8.10, 9.10</td>
</tr>
<tr>
<td>11</td>
<td>7.11, 8.11, 9.11</td>
</tr>
<tr>
<td>12</td>
<td>7.12, 8.12, 9.12</td>
</tr>
<tr>
<td>13</td>
<td>7.13, 8.13, 9.13</td>
</tr>
<tr>
<td>14</td>
<td>7.14, 8.14, 9.14</td>
</tr>
<tr>
<td>15</td>
<td>7.15, 8.15, 9.15</td>
</tr>
</tbody>
</table>

DQA1 + PolyMarker

45 possible genotypes

SNP Typing Platforms

- RT-PCR (TaqMan, Light Cycler, Molecular Beacon)
- ASPE (SNAPshot, Orchid UHT, MALDI, FP)
- Mass Spectrometry (Electrospray)
- Sequencing
- Flow Cytometry (Luminex)
- Pyrosequencing
- Ligation (SNPplex, Illumina)
- Invader assay
- ARMS assay (FSS) ASPE = allele-specific primer extension
- RFLP

Budowle 2004 FSI 130-142; Sobrino et al., 2005 FSI (epub); Dixon et al., 2005 FSI (epub)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Protocol with SNaPshot™ “Kit”

- Genomic DNA sample
- (Multiplex) PCR
- ExoSAP Digestion
- Add SNP primer(s) and SNaPshot mix
- SNP Extension (cycle sequencing)
- SAP treatment
- Sample prep for 310/3100
- Run on ABI 310/3100
- Data Analysis (GeneScan)
- Type sample (Genotyper 3.7)

 Allele-Specific Primer Extension

- SNP Primer is extended by one base unit
- “tail” used to vary electrophoretic mobility
- Oligonucleotide primer 18-28 bases
- ABI PRISM® SNaPshot™ Multiplex System
- Fluorescently labeled ddNTPs + polymerase

PCR Amplified DNA Template

- ddNTP
- Dye label
- Color
- Cycle Description
- ddNTP: dye label: color
- 25 Cycles
- 96°C 10s
- 50°C 5s
- 60°C 30s

6-plex SNP Assay Using SNaPshot

- CHR: 2 13 6 15 17
- Extension primers for 6-plex

Utility of SNP Markers

Replace Autosomal STRs?

"It is unlikely that SNPs will replace STRs as the preferred method of testing of forensic samples in the near to medium future."

Specialized applications

- mtDNA – coding region and linear arrays
- Y-SNPs – lineage, population study, sample discrimination
- Autosomal SNPs – highly degraded samples, shed hairs, physical characteristics, ethnic/geographical determination


SNP Work at NIST

- mtDNA coding region SNPs
  - Will selected SNPs aid resolution of common HV1/HV2 types?
- Y-SNPs
  - Potential for ethnicity prediction?
- Autosomal SNPs
  - Multiplexing limitations?
  - Capability of mixture detection/interpretation?
  - Performance with degraded DNA or LCN templates?

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

April 4, 2006

mtDNA Coding Region 11-plex ASPE Assay

Result from 1 pg (genomic DNA)

Now being used by Armed Forces DNA Identification Laboratory (AFDIL) for mtDNA casework

11-plex PCR and 11-plex SNP detection
Sites are polymorphic in Caucasians (H1) and useful in resolving most common HV1/HV2 types
Multiplex PCR used to co-amplify all regions of interest at once
PCR product sizes kept under 200 bp to enable success with degraded DNA samples


Forensic Utility of Y Chromosome SNPs

Y chromosome markers are useful in mixed male - female samples
Haplogroups are non-randomly distributed among populations therefore potential exists for predicting population of origin
Low mutation rate of SNPs 2 x 10^{-8} per base per generation

Typed 51 Y-SNPs using ASPE and Marligen (Luminex beads)

>250 Y-SNPs described


Typing 51 Y-SNPs

51 Y-SNPs
115 African Americans
114 Caucasians

Concordant typing results for over 19 loci
(>3,800 allele calls)


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)

Forensic Utility

51 Y-SNPs versus 1 Y-STR

For N = 211 male samples

As a stand alone forensic assay
1 Y-STR is better than 51 Y-SNPs

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

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

To date: (>100,000 allele calls)
Identifier (15 autosomal markers + Amelogenin) (10,608)
Roche Linear Arrays (HVI/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 (9,226)
New miniSTR loci—for 11 loci, 7,293 genotypes
mitDNA full control region sequences by AFIL

Autosomal SNP characteristics

• 70 Loci – sites from Orchid – C/T bi-allelic
• Present on 20 of 22 autosomal CHR (3,16,X,Y)
• Amplicon size range 59 - 108 bp (average 69)
• Markers are typed by allele-specific primer extension assays (ABI SNaPshot)
• Level of multiplexing (6-12-plexes)
• Web page for SNP site info
http://www.cstl.nist.gov/biotech/strbase/SNP.htm

6-plex SNP Assay

SNP typing results for a single individual

Allele Frequencies for 70 SNP Loci in U.S. Populations

Probability of a Random Match using 70 SNPs

Assuming all loci are unlinked

AA (N = 71)
Cauc (N = 74)
Hisp (N = 44)
for unrelated individuals

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

70 were typed for 189 U.S. samples (self-identified ethnicities)
74 Caucasians + 71 African Americans AA + 44 Hispanics

Total of 13,230 possible genotypes

42 Samples were re-injected to confirm ambiguous results (99.7% success rate on first pass)


“Best” 12 loci combined into a 12plex SNP assay

Vallone et al. (2006) Progress in Forensic Genetics, in press

29plex SNP Assay Developed by SNPforID (EDNAP Study Results Generated at NIST)

SNP 29plex Result (Color Separated)
NIST Work Conclusions

- mtSNPs: Coding region SNPs can fulfill a useful role for separating common HV1/HV2 mitotypes.
- Y-SNPs: Y-SNPs will have limited utility for individualizing a sample. Determination of ethnic origin may be challenging for U.S. samples.
- Autosomal SNPs: 12plex assay shows some promise for typing degraded samples and shed hairs. Would be used in conjunction with STR kits rather than in place of them.

NIST Work with SNP Loci

- mtDNA coding region SNP 11plex assay
- U.S. population information with 50 Y-SNPs
- U.S. population frequencies with 70 autosomal SNPs
- Construction of 12plex autosomal SNP assay
  - Vallone et al. (2006) Progress in Forensic Genetics 11
- Creation of Forensic SNP Information website on STRBase

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

(Selected) SNP Work by Others

- SNPforID (http://www.snpforid.org) – team of five European labs working on selection of useful autosomal SNP loci, developing multiplex assays, and collecting population data
- Manfred Kayser’s group – seeking minimal set of SNP loci for distinguishing ethnicities
  - AJHG 2006 78: 680-690
- Ken Kidd’s group – seeking “best” forensic SNPs
  - FSI article, in press

Comparison of STRs and SNPs

- Conventional STR
  - Larger target region (miniSTR targets same region)
  - More possible variants than SNPs
  - Only need a moderate number of STR markers
  - Range of sizes examined (e.g., 28 bp spread if 4 bp repeat)
  - Mixtures easier to detect and decipher

- SNP
  - Smaller target region
  - Fewer possible variants
  - Need more SNP markers
  - Constant size examined

STR Markers Ability to Detect DNA Mixtures

Green channel from Identifier STR tests

Well A4

Well A5

All 5 loci shown exhibit 3 alleles suggesting that a DNA mixture from at least two individuals is present

STRs permit fairly easy mixture identification due to the number of possible alleles and the high heterozygosities of loci used for human identity testing

SNP 6plex with the Same DNA Samples

SNP Locus 1 2 3 4 5 6

Well A4

Single source DNA

Well A5

Mixed DNA (according to STR data)

Mixture in well A5 would probably go undetected with this 6plex SNP panel (except for heterozygote peak imbalance at a single locus)

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
A miniSTR is a reduced size STR amplicon that enables higher recovery of information from degraded DNA samples. Testing must be performed to show allele concordance between primer sets.

Comparison of PCR Amplification Success Rates with Commercial Kit vs. miniSTR Assays

Study with 31 human bones from the "Body Farm" (Knoxville, TN) and Franklin County Coroner’s Office (OH)

Comparison with commercial kits:
- Better success with degraded DNA
- Better success with low amounts of DNA
- Better capacity for handling mixed DNA samples
- Concordance to STR loci in commercial kits is possible

Characterization of New miniSTR Loci

Candidate STR marker selection

Pull-down sequence data from the web

Analytically: Genotype Location

Screen for PCR Primers

Test primers for Multiplex-ability

"Laboratory work"

Test Markers on Reputable samples

Separate Heterozygous to observe allele sizes

Build Models for Genotyping

Characterized Allele Ladders

"Computer work"

Characterization of New STRs

Disadvantages
- Not all commonly used STRs can be made significantly smaller—thus new loci will be needed
- Cannot multiplex as many loci due to size constraints
- No commercial kit (yet)
- STR flanking region mutations may make results disconcordant (e.g., D13 and VWA deletions)

Initial Testing Results with Potential miniSTR Loci


21 additional loci now examined

27 new miniSTRs have been characterized

http://www.cstl.nist.gov/biotech/strbase/newSTRs.htm
### Miniplex "NC01"

**PCR Product Size (bp)**

- **EF3A** (blue) 103, 112, 132, 142
- **D16S539** 85, 98
- **D10S1248** 112, 132
- **D22S1045** 106, 114
- **D14S1434** 106, 114
- **D22S1045** 106, 114
- **VIC** (green) 103, 112, 132, 142
- **D16S539** 85, 98
- **D10S1248** 112, 132
- **D22S1045** 106, 114
- **NED** (yellow) 103, 112, 132, 142
- **D16S539** 85, 98
- **D10S1248** 112, 132
- **D22S1045** 106, 114

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

---

### Direct Comparisons of SNPs and STRs on Degraded DNA Templates

- **World Trade Center DNA Investigation**
  - A panel of 70 SNPs run on >15,000 bone extracts by Orchid Cellmark; no additional identifications made
  - Reduced size STR markers (miniSTRs) aided last 20% of WTC DNA identifications

- **EDNAP Degraded DNA Study**
  - Organized by Lindsey Dixon and Peter Gill (UK FSS); involved 9 labs testing artificially degraded blood and saliva stains
  - miniSTRs outperformed SNPs

---

### EDNAP Exercise on Degraded DNA

**ARTICLE IN PRESS**

- Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (EDNAP) exercise
  - L.A. Dixon a, A.E. Dobbinb, H.K. Parkerc, J.M. Butler a, P.M. Vallsone

**MiniSTR primer mixes and allelic ladders were provided by NIST**

---

### Recent Article Advocating miniSTRs

They recommend that miniSTRs “be adopted as the way forward to increase both the robustness and sensitivity of analysis.”

**They recommend that European laboratories adopt three new mini-STR loci, namely: D10S1248, D14S1434 and D22S1045. (D14 now replaced by D2S441)**

---

### Comparison of STR Locus Variability

<table>
<thead>
<tr>
<th>Locus</th>
<th>N</th>
<th>Heterozygosity</th>
<th>Rank</th>
<th>Size Range (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D10S1248</td>
<td>669</td>
<td>0.987</td>
<td>2</td>
<td>127 - 139 (CoFiler)</td>
</tr>
<tr>
<td>D16S539</td>
<td>681</td>
<td>0.874</td>
<td>14</td>
<td>81 - 142 (CoFiler)</td>
</tr>
<tr>
<td>D18S51</td>
<td>659</td>
<td>0.970</td>
<td>1</td>
<td>81 - 142 (CoFiler)</td>
</tr>
<tr>
<td>D9S2157</td>
<td>660</td>
<td>0.831</td>
<td>12</td>
<td>100 - 110 (CoFiler)</td>
</tr>
<tr>
<td>D21S11</td>
<td>659</td>
<td>0.866</td>
<td>6</td>
<td>41 - 49 (CoFiler)</td>
</tr>
<tr>
<td>D12S1082</td>
<td>681</td>
<td>0.976</td>
<td>5</td>
<td>82 - 139 (CoFiler)</td>
</tr>
<tr>
<td>D14S1434</td>
<td>663</td>
<td>0.970</td>
<td>3</td>
<td>123 - 159 (CoFiler)</td>
</tr>
<tr>
<td>D19S433</td>
<td>659</td>
<td>0.969</td>
<td>11</td>
<td>139 - 157 (CoFiler)</td>
</tr>
</tbody>
</table>

**EDNAP/ENFSI suggested markers**

- <150 bp

---

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

April 4, 2006

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

Chromosomal Locations for New miniSTRs

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Location</th>
<th>Number</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>D10S1248</td>
<td>NC01</td>
<td>164</td>
<td>African Americans</td>
</tr>
<tr>
<td>D17S1301</td>
<td>NC01</td>
<td>170</td>
<td>U.S. Caucasians</td>
</tr>
<tr>
<td>D11S4463</td>
<td>NC01</td>
<td>140</td>
<td>U.S. Hispanics</td>
</tr>
<tr>
<td>D19S433</td>
<td>NC02</td>
<td>142</td>
<td>Japanese</td>
</tr>
<tr>
<td>D2S1338</td>
<td>NC02</td>
<td>185</td>
<td>Chinese</td>
</tr>
<tr>
<td>D3S1358</td>
<td>NC02</td>
<td>182</td>
<td>Malaysian</td>
</tr>
<tr>
<td>D5S818</td>
<td>NC02</td>
<td>178</td>
<td>East Indian</td>
</tr>
<tr>
<td>TPOX</td>
<td>NC02</td>
<td>100</td>
<td>Italian</td>
</tr>
<tr>
<td>VWA</td>
<td>NC02</td>
<td>200</td>
<td>Romanian</td>
</tr>
<tr>
<td>TH01</td>
<td>NC02</td>
<td>226</td>
<td>Chinese</td>
</tr>
<tr>
<td>Penta D</td>
<td>NC02</td>
<td>238</td>
<td>Malaysian</td>
</tr>
<tr>
<td>Penta E</td>
<td>NC02</td>
<td>238</td>
<td>East Indian</td>
</tr>
</tbody>
</table>

Source: Coble and Butler (2005)

New miniSGM miniplex: AMEL, TH01, FGA, D18, D16, D2

EDNAP/ENFSI degraded DNA study coordinated by Peter Gill

Creation of miniSTR information on STRBase

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

miniSTRs for Degraded DNA

- Original miniSTR paper with CODIS loci, D2, D19, Penta D, Penta E
- Many CODIS loci are too big and make poor miniSTRs
- New miniSTRs and assays: NC01, NC02
- New miniSGM miniplex: AMEL, TH01, FGA, D18, D16, D2
- EDNAP/ENFSI degraded DNA study coordinated by Peter Gill
- Creation of miniSTR information on STRBase

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

Score Card

<table>
<thead>
<tr>
<th>STRs/miniSTRs</th>
<th>SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success with Degraded DNA</td>
<td>✔</td>
</tr>
<tr>
<td>Power of Discrimination</td>
<td>✔</td>
</tr>
<tr>
<td>Mixture Det./ Interpretation</td>
<td>✔</td>
</tr>
<tr>
<td>Other Applications: Ethnicity Estimation, Physical Traits, etc.</td>
<td>✗</td>
</tr>
</tbody>
</table>

Acknowledgments

Funding from interagency agreement 2003-IJ-R-029 between NIJ and the NIST Office of Law Enforcement Standards

Past and Present Collaborators (also funded by NIJ):
- Mike Coble and Alan Redd (U. AZ) for Y-chromosome studies
- Tom Parsons, Rebecca Just, Joel Frein (AFDIL) for mtDNA coding SNP work
- Sandy Calloway (Roche) for mtDNA LINEAR ARRAYs
- Bruce McCord and students (FL Int. U.) for miniSTR work
- Marilyn Raymond and Victor David (NCI-Frederick) for cat STR work
- Artie Eisenberg and John Plant (U. North Texas) for miniSTR testing on bones
- Murray Brilliant (U. AZ) for phenotype markers
- Ken Kidd (Yale U.) for SNP typing population samples

Past and Present Collaborators:

John Butler
Margaret Kline
Peter Valente
Mike Coble
Jon Redman
Amy Decker
Becky Hill
Chris DeAngelis
Dave Duewer

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

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