STR and Molecular Biology Artifacts

Dr. John M. Butler
National Institute of Standards and Technology
john.butler@nist.gov

Types of STR Repeat Units

- **Di-nucleotide** (CA)(CA)(CA)(CA)
- **Tri-nucleotide** (GCC)(GCC)(GCC)
- **Tetra-nucleotide** (AATG)(AATG)(AATG)
- **Penta-nucleotide** (AGAAA)(AGAAA)
- **Hexa-nucleotide** (AGTACA)(AGTACA)

Short tandem repeat (STR) = microsatellite = simple sequence repeat (SSR)

How many STRs in the human genome?

- The efforts of the Human Genome Project have increased knowledge regarding the human genome, and hence there are many more STR loci available now than there were 10 years ago when the 13 CODIS core loci were selected.
- More than 20,000 tetranucleotide STR loci have been characterized in the human genome (Collins et al. An exhaustive DNA microsatellite map of the human genome using high performance computing. Genomics 2003;82:10-19).
- There may be more than a million STR loci present depending on how they are counted (Ellegren H. Microsatellites: simple sequences with complex evolution. Nature Rev Genet 2004;5:435-445).

Categories for STR Markers

<table>
<thead>
<tr>
<th>Category</th>
<th>Example Repeat Structure</th>
<th>13 CODIS Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple repeats – contain units of identical length and sequence</td>
<td>(GATA)(GATA)(GATA)</td>
<td>TPOX, CSF1PO, DSS818, D13S317, D16S539</td>
</tr>
<tr>
<td>Simple repeats with non-consensus alleles (e.g., TH01 9:3)</td>
<td>(GATA)(CAT)-(GATA)</td>
<td>TH01, D18S51, DTS620</td>
</tr>
<tr>
<td>Compound repeats – comprise two or more adjacent simple repeats</td>
<td>(GATA)(GATA)(GACA)</td>
<td>WFA, FGA, D3S1358, D8S1179</td>
</tr>
<tr>
<td>Complex repeats – contain several repeat blocks of variable unit length</td>
<td>(GATA)(GACA)(CA)(CATA)</td>
<td>D2S1911</td>
</tr>
</tbody>
</table>

These categories were first described by Urquhart et al. (1994) Int. J. Legal Med. 107:13-20

Advantages of Multiplex PCR

- **Compatible primers are the key to successful multiplex PCR**
- STR kits are commercially available
- 15 or more STR loci can be simultaneously amplified

**Challenges to Multiplexing**
- Primer design to find compatible primers (no program exists)
- Reaction optimization is highly empirical often taking months

Advantages of Multiplex PCR

- Increases information obtained per unit time (increases power of discrimination)
- Reduces labor to obtain results
- Reduces template required (smaller sample consumed)


http://www.cstl.nist.gov/biotech/strbase/training.htm
Biological “Artifacts” of STR Markers

- Stutter Products
- Non-template nucleotide addition
- Microvariants
- Tri-allelic patterns
- Null alleles
- Mutations

Chapter 6 covers these topics in detail

Stutter Products

- Peaks that show up primarily one repeat less than the true allele as a result of strand slippage during DNA synthesis
- Stutter is less pronounced with larger repeat unit sizes (dinucleotides > tri- > tetra- > penta-)
- Longer repeat regions generate more stutter
- Each successive stutter product is less intense (allele > repeat-1 > repeat-2)
- Stutter peaks make mixture analysis more difficult

Stutter Product Formation

Repeat unit bulges out when strand breathing occurs during replication

Typically 5-15% of true allele in tetranucleotide repeats STR loci

Occurs less frequently (typically <2%) – often down in the “noise” depending on sensitivity

Deletion caused by slippage on the copied (bottom) strand

Insertion caused by slippage of the copying (top) strand
Non-Template Addition

- Taq polymerase will often add an extra nucleotide to the end of a PCR product; most often an “A” (termed “adenylation”)
- Dependent on 5’-end of the reverse primer; a “G” can be put at the end of a primer to promote non-template addition
- Can be enhanced with extension soak at the end of the PCR cycle (e.g., 15-45 min @ 60 or 72 °C) – to give polymerase more time
- Excess amounts of DNA template in the PCR reaction can result in incomplete adenylation (not enough polymerase to go around)
- Best if there is NOT a mixture of “+/- A” peaks (desirable to have full adenylation to avoid split peaks)

Impact of the 5’ Nucleotide on Non-Template Addition

- Last Base for Primer Opposite Dye Label
  - PCR conditions are the same for these two samples
- Promega includes an ATT sequence on the 5'-end of many of their unlabeled PP16 primers to promote adenylation

Higher Levels of DNA Lead to Incomplete Adenylation

- DNA Size (bp)
- Relative Fluorescence (RFUs)
- Stochastic effect when amplifying low levels of DNA produces allele dropout

Impact of DNA Amount into PCR

- Reason that DNA Quantitation is Important Prior to Multiplex Amplification
- Generally 0.5 - 2.0 ng DNA template is best for STR kits
  - Too much DNA
    - Off-scale peaks
    - Split peaks (+/-A)
    - Locus-to-locus imbalance
  - Too little DNA
    - Heterozygote peak imbalance
    - Allele drop-out
    - Locus-to-locus imbalance

Microvariant “Off-Ladder” Alleles

- Defined as alleles that are not exact multiples of the basic repeat motif or sequence variants of the repeat motif or both
- Alleles with partial repeat units are designated by the number of full repeats and then a decimal point followed by the number of bases in the partial repeat (Bar et al. Int. J. Legal Med. 1994, 107:159-160)
- Example: TH01 9.3 allele: [TCAT]_{4.5}–CAT [TCAT]_{5}

http://www.cstl.nist.gov/biotech/strbase/training.htm
An Example of an “Off-Ladder” Microvariant at the Yfiler Locus DYS635

Variant Alleles Cataloged in STRBase

Null Alleles

Concordance between STR primer sets is important for DNA databases

vWA Primer Position Comparisons

http://www.cstl.nist.gov/biotech/strbase/training.htm
Apparent Null Alleles Observed During Concordance Studies

<table>
<thead>
<tr>
<th>STR Locus</th>
<th>New Section of STRBase</th>
<th>Allele drops out</th>
<th>Allele 6 amplicon has &quot;dropped out&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR Kit/Assays</td>
<td>Impact of DNA Sequence Variation in the PCR Primer Binding Site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D21S11</td>
<td>Shift of alleles 10 and 11 due to deletion outside of miniplexes away from primer binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D5S818, D7S820</td>
<td>Mutation in middle of primer binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D16S539</td>
<td>Mutation at 3'-end of primer binding site (allele dropout)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

New Section of STRBase (launched to track MiniFiler discordance and allele dropout frequency):

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

<table>
<thead>
<tr>
<th>STR Locus</th>
<th>STR Measured Mutation Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR Kit/Assays</td>
<td>summary of STR Mutations</td>
</tr>
<tr>
<td>13 CODIS loci affected so far</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Impact of DNA Sequence Variation in the PCR Primer Binding Site</th>
</tr>
</thead>
</table>

Mutation Observed in Family Trio

<table>
<thead>
<tr>
<th>STR Locus</th>
<th>STR Measured Mutation Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR Kit/Assays</td>
<td>Summary of STR Mutations</td>
</tr>
<tr>
<td>13 CODIS loci affected so far</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Impact of DNA Sequence Variation in the PCR Primer Binding Site</th>
</tr>
</thead>
</table>

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

D18S51 Null Allele from Kuwait Samples with ABI Primers


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

Mutation at 3'-end of primer binding site (allele dropout)
General Information
• Intro to STRs (downloadable PowerPoint)
• STR Fact Sheets
• Sequence Information
• Multiplex STR Kits
• Variant Allele Reports
• Training Slides

Forensic Interest Data
• FBI CODIS Core Loci
• DAB Standards
• NIST SRMs 2391
• Published PCR Primers
• Y-Chromosome STRs
• Population Data
• Validation Studies
• miniSTRs

Supplemental Info
• Reference List
• Technology Review
• Addresses for Scientists
• Links to Other Web Sites
• DNA Quantitation
• mtDNA
• New STRs

New information is added regularly…