Commonly Used STR Markers
Repeats

• Satellites
  – 100 to 1000 bases repeated

• Minisatellites
  – VNTR – variable number tandem repeat
  – 10 to 100 bases repeated

• Microsatellites
  – STR – short tandem repeat
  – 2 to 6 bases repeated
  – Most commonly used for Forensics
Advantages of STRs

• Occur frequently in genome
  – About every 10,000 bases or so
• Easily amplified by PCR
• Both alleles are very similar in size
  – Don’t have problems with allele drop out
  – Many markers can be multiplexed
• Highly variable polymorphisms
• Large number of STRs have been characterized and studied
Types of STRs

Length of repeat:
• Dinucleotide
  – 2 bases repeated – ex. AC
• Tri-
  – 3 bases repeated – ex. AAC
• Tetra-
  – 4 bases repeated – ex. AGAT
• Penta- and Hexa-
Types of STRs

Pattern of the repeat:

- **Simple:**
  - Repeat units all identical length and sequence
- **Compound:**
  - Two or more adjacent simple repeats
- **Complex:**
  - Repeats of variable length or sequences
- **Others**
Microvariants

• Allele that has an incomplete repeat unit
• Can happen even with simple repeats
• Example:
  – Allele 9.3 of TH01 locus
  – Has 9 copies of 4 base pair repeat
  – Plus 1 copy that only contains 3 of the 4
  – Must be validated
  – Otherwise may just be an error in genotyping methodology – not actually a variant
Desirable STR Markers

• Highest possible variation:
  – Each marker gives a lot of information
• Ability to genotype using small PCR product (less than 400 bps):
  – Can work with degraded DNA sample
• Less “stuttering”
• Ability to resolve all alleles clearly:
  – Easier to resolve 4 base pair difference
  – Rather than 3, 3 better than 2, etc
Selecting Candidate STRs

- High discriminating power:
  - Heterozygosity > 70%
  - More than 70% of individuals will be hetero.
- Separate chromosomes:
  - Loci are unlinked genetically
- Reproducibility of genotypes
- Low mutation rates
- Low stutter rates
- Smaller PCR product sizes
Common Nomenclature

Must agree how to name and genotype STRs so that matches can be made

• Name is based on core repeat unit:
  – Example – AGAT

• Also need to agree on where to start counting number of repeats:
  – First time repeat unit appears = 1

• Both of these will change depending on which strand of DNA is used
Depends on which strand

5’-TTTCCC TCAT TCAT TCAT TCAT TCAT TCAT TCACCATGGA-3’
3’-AAAGGG AGTA AGTA AGTA AGTA AGTA AGTA AGTGGTACCT-5’
DNA Commission Guidelines

1. For markers in genes – always use coding strand
2. For D#S### markers – always use first published report of marker
3. First repeat is the first 5’ nucleotides that define repeat unit
4. Microvariants:
   • Number of complete repeats, decimal point, then number of bases in incomplete repeat
Ladders

• Within every lane of gel need a ladder to accurately identify allele sizes
• Must spread across any size range that alleles might contain
• Made from STR
  – Find representative alleles that span all population variants
• Commercially available
13 CODIS Markers

COmbined DNA Index System - USA

• Agreed upon 13 “core” STR loci
• When all 13 are tested probability of random match:
  – Less than 1 in 1 Trillion
  – Only 6 Billion people on earth
• Kits are available to genotype all 13 markers with less than 1 ng of DNA in a few hours
13 Markers

• On 12 different chromosomes
• All autosomal:
  – Which means gender is not genotyped as one of 13 core loci
  – Use marker AMEL for gender
• Types of markers:
  – Simple repeats (with or without microvariants)
  – Compound repeats
  – Complex repeats
Commercial Kits

- All primers are designed and validated
- Markers are multiplexed and optimized
- Ladder is included in every reaction
- Come with positive control DNA
- Saves Forensic laboratories time and effort of all this optimization
- More confidence in sharing data
- Genotyping data gains confidence in court
Linkers

- A linker can be added to end of primer
- Then linker will become part of PCR product
- Linker used to shift the mobility of product through the gel
- This allows two markers that overlap to be genotyped simultaneously because one will be shifted away from other
(A) COfiler kit
allele relative size ranges

Size overlap

6  D7S820  15
256.01 bp  292.62 bp

NED-labeled (yellow)

279.65 bp  317.67 bp

CSF1PO

J OE-labeled (green)

10 non-nucleotide linkers = ~ +25 bp shift

(B) Identifiler kit
allele relative size ranges

6  D7S820  15
255.15 bp  291.58 bp

6FAM-labeled (blue)

291.58 bp  304.69 bp

10 non-nucleotide linkers = ~ +25 bp shift

6  CSF1PO  15
304.69 bp  341.84 bp

6FAM-labeled (blue)

Figure 5.8, J.M. Butler (2005) Forensic DNA Typing, 2nd Edition © 2005 Elsevier Academic Press
Linkers

Example -

• D7S820 – 256 to 292 bps product
• CSF1P0 – 279 to 317 bps product
• Can either label them with two different colors:
  – D7S820 = yellow, CSF1P0 = green
• Or add a linker to one marker:
  – CSF1P0 is shifted to produce 304 to 341 bps product
Specific Markers

• Going to go through some details about the 13 core loci
• Plus the gender loci everyone uses AMEL
• Should know what chromosome each STR marker is on
• Any other details you don’t need to memorize
• Be able to read D#S### marker names
CSF1P0

- 5q33.1
- Chromosome 5 – around 149 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat motif:
  - TAGA
- 6th intron of proto-oncogene c-fms
FGA

- 4q31.3
- Chromosome 4 – around 156 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat Motif:
  - CTTT
- 3rd intron of alpha fibrinogen gene
TH01

- 11p15.5
- Chromosome 11 – around 2 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat Motif:
  - TCAT
- 1\textsuperscript{st} intron of tyrosine hydroxylase gene
TPOX

- 2p25.3
- Chromosome 2 – around 1 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat Motif:
  - GAAT
- 10th intron of thyroid peroxidase gene
VWA

- 12p13.31
- Chromosome 12 – around 20 Mb
- Tetranucleotide repeat
- Compound STR
- Repeat Motif:
  - [TCTG][TCTA]
- 40\textsuperscript{th} intron of von Willebrand Factor gene
D3S1358

- 3p21.31
- Chromosome 3 – around 45 Mb
- Tetranucleotide repeat
- Compound STR
- Repeat Motif:
  - [TCTG][TCTA]
- Not related to any gene
D5S818

- 5q23.2
- Chromosome 5 – around 123 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat Motif:
  - AGAT
- Not related to any gene
D7S820

- 7q21.11
- Chromosome 7 – around 83 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat Motif:
  - GATA
- Not related to any gene
D8S1179

- 8q24.13
- Chromosome 8 – around 125 Mb
- Tetranucleotide repeat
- Compound STR
- Repeat Motif:
  - [TCTA] [TCTG]
- Not related to any gene
D13S317

- 13q31.1
- Chromosome 13 – around 80 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat Motif:
  - TATC
- Not related to any gene
D16S539

• 16q24.1
• Chromosome 16 – around 86 Mb
• Tetranucleotide repeat
• Simple STR
• Repeat Motif:
  – GATA
• Not related to any gene
D18S51

- 18q21.33
- Chromosome 18 – around 59 Mb
- Tetranucleotide repeat
- Simple STR
- Repeat Motif:
  - AGAA
- Not related to any gene
D21S11

- 21q21.1
- Chromosome 21 – around 19 Mb
- Tetranucleotide repeat
- Complex STR
- Repeat Motif:
  - [TCTA][TCTG] surrounded by a constant section of specific sequence
- Not related to any gene
Amelogenin

- Amelogenin gene encodes for protein in tooth enamel
- Gene on X chromosome
- But also on the part of the Y chromosome that is homologous to X chromosome:
  - PseudoAutosomal Region (PAR)
- Therefore this gene is actually on both X and Y chromosomes
Amelogenin

- AMEL loci
- Primers are homologous to one region on both X and Y chromosome
- X chromosome has 6 bp deletion and Y chromosome doesn’t
- Therefore XX genotype will be homozygous – identify females
- XY genotype will be heterozygous – identify males
Y STRs

• Other STRs exist only on Y chromosome

• Excellent for separating male and female mixed samples

• What is advantage to using AMEL over using a Y chromosome STR?
• What is advantage of using Y STR?
Specific Markers

- There are additional markers that are commonly used for Forensic other than these 14
- We aren’t going to go over any others
- Should know what chromosome each STR marker is on
- Any other details you don’t need to memorize
STRBase

- References for all STRs
- Provides agreed upon strand and repeat motif information
- Primer sequences
- List common alleles and allele frequencies
- Any microvariants
- Success rates and scientists names
Any Questions?

Read Chapter Six
Commercially Available Kits:

Promega Corporation
• 13 core loci + 2 additional loci
• Plus AMEL for gender
• 4 dyes total (3 for alleles and 1 for ladder)

Applied Biosystems
• 13 core loci + 2 additional loci
• Plus AMEL for gender
• 5 dyes total (4 for alleles and 1 for ladder)