Short Tandem Repeat (STR) Analysis

Maj Gen (R) Suhaib Ahmed, HI (M)

Short tandem repeats (STR) are randomly distributed DNA sequences in which 2-6bp are tandemly repeated. These are scattered on all chromosomes including the autosomes as well as the sex chromosomes. An example of four base pair (tetranucleotide) repeat STR on chromosome 5 is shown in Fig. 9.1 The sequence "GATA" is repeated several times. The numbers of STR repeats vary between individuals and the alleles are named according to the number of repeat units these may contain. Every individual inherits an allele each from its parents. The two alleles may be identical (homozygote) or different (heterozygote or more precisely compound heterozygote).

ATATTTTAATAGCAAGTATGTGACAAGGGTGATTTTCCTCTTTGGTATCCTTATGTA
ATATTTTGAAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGGT
AGATAGAGGTATAAATAAGGATACAGATATAGNTACAAATGTTGTAAACTGTGGCTA
TGATTGGAATCACTTGGCTAAAAAGCGCTNAAGCNTTCCTCTGNGAGAGGCAATTAC

Fig. 9.1 GATA repeats at the D5S818 locus (GenBank: G08446.1) The sequence GATA is repeated 11 times.

PCR amplification of STRs

STRs can be amplified by a pair of primers flanking the tandem repeats. The resulting amplified products vary in length depending on the number of repeat units present in the allele (Fig. 9.2). In a compound heterozygote with different number of repeats on the two chromosomes, two different sized amplified products are formed. In a homozygote with the same number of repeats on the two chromosomes two amplified fragments of the same size are produced. The amplified fragments vary in length depending on the number of repeats in the allele. If at a tetranucleotide repeat locus one allele has 5 repeats of 4bp each and the other allele has 8 repeats the two amplified fragments would differ by 12bp (4 + 4 + 4).

Polyacrylamide gel electrophoresis

The amplified fragments of STRs at any given locus are only a few base pairs different. Such small differences are best resolved by Polyacrylamide gel electrophoresis. After PCR amplification the STRs are analysed by relatively long polyacrylamide gels. The gels are usually run for several hours. Various alleles are recognized by simultaneously running accurate DNA size standards or allelic ladders (Fig. 9.2B).

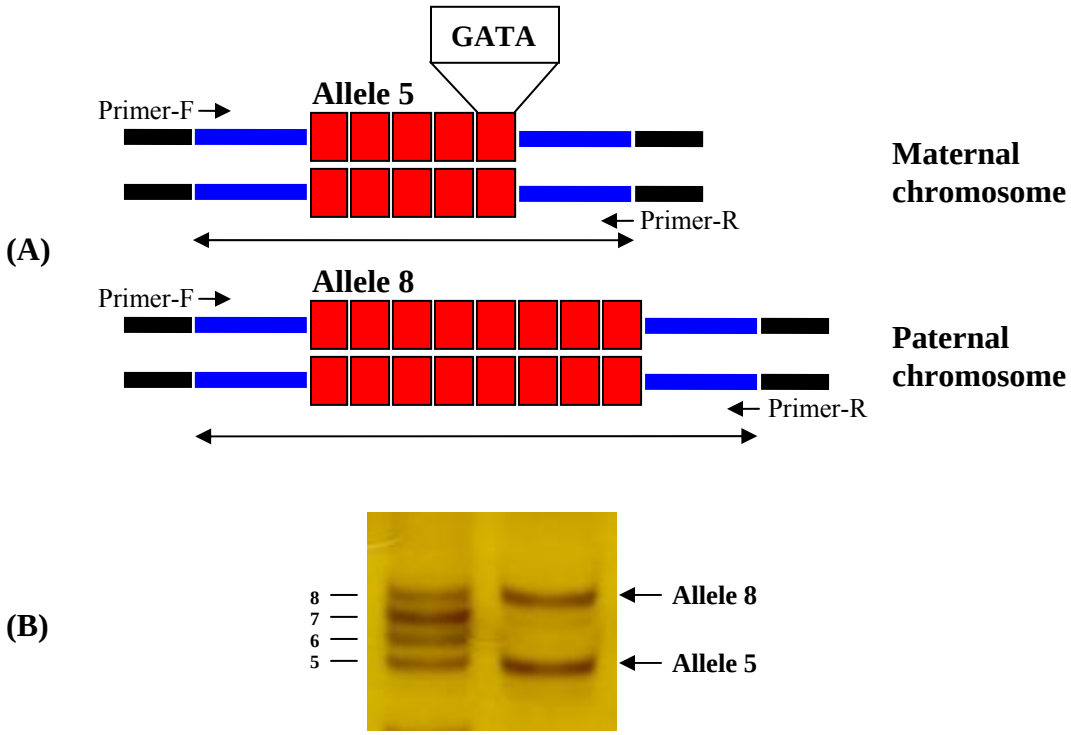

Fig. 9.2. (A) PCR amplification of STR locus D5S818 having "GATA" repeats. The tandem repeats on the maternal and the paternal chromosomes are amplified by a pair of primers (F & R) flanking the repeat sequences. (B) Polyacrylamide gel electrophoresis of the amplified product and allelic ladder shows the presence of allele 5 and 8.

Genetic analyzer

The manual sizing of STR alleles is laborious and time consuming. Automated genetic analyzer provides an extremely efficient but expensive alternate to the manual method. It uses a long thin bored capillary for electrophoresis of amplified DNA (Chapter 5). The STRs are amplified by fluorescent labeled primers and are run with a fluorescent labeled DNA size standard in the same capillary. The genetic analyzer detects the amplified products by laser excitation of the fluorescent dyes. It is an extremely sensitive and accurate method of STR allele sizing. Very small amounts of target DNA (<0.5ng) can be detected. Its other advantages include high speed and analysis of multiple amplified products labeled with different fluorescent dyes in the same tube. Commercial kits with amplification of up to fifteen different STR loci in the same tube are also available.

STR analysis by manual method

Locus:	D5S818 (Fig 9.3)
GenBank accession:	G08446.1

Forward primer: 5'-AAGGGTGATTTTCCTCTTTGG
Reverse primer: 5'- AGCGCTTTTTAGCCAAGTGA
Amplified product: 136-176 bp (1-11 repeats)

TCTAATTAAAGTGGTGTCCCAGATAATCTGTACTAATAAAAGTATATTTTAATAGC
AAGTATGTGACAAGGGTGATTTTCCTCTTTGGTATCCTTATGTAATATTTTGAAGA
TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGGTAGATAGAGGTA
TAAATAAGGATACAGATATAGNTACAAATGTTGTAAACTGTGGCTATGATTGGAAT
CACTTGGCTAAAAAGCGCTNAAGCNTTCCTCTGNGAGAGGCAATTACTTTTTTNCT
TAGGNACTNCCTCANCAGTCTNTTNGC

Fig. 9.3. PCR primers for amplification of the GATA repeat at D5S818 locus (GenBank accession: G08446.1).

Reaction conditions

- Reaction volume: 25 µl
- Primer concentration: 1 µl (5 pmol each/µl)
- Taq polymerase: 0.5 units (0.1µl)
- Template DNA: 1µl (200ng)
- Thermal cycling:
  - Initial denaturation: 1 minute at 94°C
  - No. of cycles: 30
    Denaturation: 48 seconds at 93°C
    Annealing: 48 seconds at 60°C
    Extension: 1 minute at 72°C
    Final extension: 3 minute at 72°C
- Electrophoresis: 16 X 20 X 0.1cm 6% polyacrylamide gel.
  Load 3µl amplified product in 3µl loading dye
  Include 3µl allelic ladder in 3µl loading dye
  Run for 5 hours at 200 volts.
- Staining: 0.1% Silver nitrate.
- Result: Fig 9.4

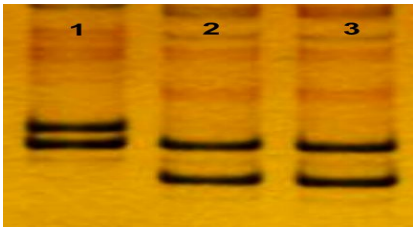

Fig. 9.4. Polyacrylamide gel electrophoresis of the PCR products of amplification at D5S818 locus.

Preparation of allelic ladder for STR analysis by gel electrophoresis

1. Select DNA samples of known genotypes at different loci previously analyzed on genetic analyzer. The selected samples should represent the alleles usually found in the target population.
2. Amplify each DNA sample in multiples of five at the locus for which the ladder is required. Use the amplification protocol described above.
3. Pool the multiply amplified products of all the samples.
4. The amplified DNA is a potent source of cross contamination therefore only use pipettes marked as "amplified DNA only".
5. Mix the pooled amplified products and make aliquots of 100µl.
6. Store at -20^{o}C.
7. Use 3µl of the pooled amplified products (allelic ladder) for each electrophoresis run (Fig. 9.4).

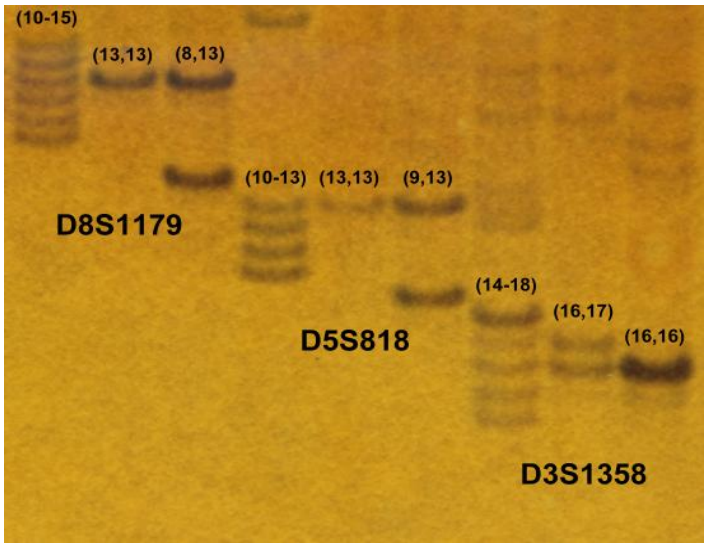

Fig. 9.4 Silver stained polyacrylamide gel electrophoresis of PCR amplified products at D8S1179, D5S818 and D3S1358 loci and the allelic ladders. The allele sizes in each allelic ladder and two unknown samples are shown in brackets.

PCR primers for amplification of STR loci

The lists of primers for amplification of autosomal and Y-chromosomal STRs are given in Table 9.1 & 9.2.

Table 9.1. PCR primers for CODIS core STR Loci.

Locus	Primer sequence
D3S1358-F	5'-ACTGCAGTCCAATCTGGGT
D3S1358-R	5'-ATGAAATCAACAGAGGCTTG
D5S818-F	5'-GGGTGATTTTCCTCTTTGGT
D5S818-R	5'-TGATTCCAATCATAGCCACA
D7S820-F	5'-TGTCATAGTTTAGAACGAACTAACG
D7S820-R	5'-CTGAGGTATCAAAAACTCAGAGG
D8S1179-F	5'-TTTTTGTATTTCATGTGTACATTCG
D8S1179-R	5'-CGTAGCTATAATTAGTTCATTTTCA
D13S317-F	5'-ACAGAAGTCTGGGATGTGGA
D13S317-R	5'-GCCCAAAAAGACAGACAGAA
D16S539-F	5'-GATCCCAAGCTCTTCCTCTT
D16S539-R	5'-ACGTTTGTGTGTGCATCTGT
D18S51-F	5'-CAAACCCGACTACCAGCAAC
D18S51-R	5'-GAGCCATGTTCATGCCACTG
FGA-F	5'-GCCCCATAGGTTTTGAACTCA
FGA-R	5'-TGATTTGTCTGTAATTGCCAGC
TPOX-F	5'-CACTAGCACCCAGAACCGTC
TPOX-R	5'-CCTTGTCAGCGTTTATTTGCC
CSF1PO-F	5'-AACCTGAGTCTGCCAAGGACTAGC
CSF1PO-R	5'-TTCCACACACCACTGGCCATCTTC
TH01-F	5'-GTGGGCTGAAAAGCTCCCGATTAT
TH01-R	5'-ATTCAAAGGGTATCTGGGCTCTGG
vWA-F	5'-CCCTAGTGGATAAGAATAATC
vWA-R	5'-GGACAGATGATAAATACATAGGATGGATGG
Amgl-F	5'-ACCTCATCCTGGGCACCCTGG
Amgl-R	5'-AGGCTTGAGGCCAACCATCAG

Table: 9.2. PCR primers for Y-STR loci.

Locus	Primer sequence
DYS19-F	5'-CTACTGAGTTTCTGTTATAGT
DYS19-R	5'-ATGGCCATGTAGTGAGGACA
DYS385a/b-F	5'-AGCATGGGTGACAGAGCTA
DYS385a/b-R	5'-TGGGATGCTAGGTAAAGCTG
DYS389-I/II-F	5'-CCAACTCTCATCTGTATTATCTAT
DYS389-I/II-R	5'-TCTTATCTCCACCCAGA
DYS390-F	5'-TATATTTTACACATTTTTGGGCC
DYS390-R	5'-TGACAGTAAAATGAACACATTGC
DYS391-F	5'-CTATTCATTCAATCATACACCCA
DYS391-R	5'-GATTCTTTGTGGTGGGTCTG
DYS393-F	5'-GTGGTCTTCTACTTGTGTCAATAC
DYS393-R	5'-AACTCAAGTCCAAAAAATGAGG
DYS438-F	5'-TGGGGAATAGTTGAACGGTAA
DYS438-R	5'-GTGGCAGACGCCTATAATCC

STR analysis by Genetic analyzer

A method for multiplex analysis of three STR loci (D13S631, D18S51 and D21S11) is described. The primers for the three loci are described in Fig. 9.5. One primer of each pair is labeled with a fluorescent dye. Since the amplified products of the three loci do not overlap the three primers are labeled with the same fluorescent dye (FAM).

(A) D13S317 (GenBank: G09017.1)

TGGGATGGGTTGCTGGACATGGTATCACAGAAGTCTGGGATGTGGAGGAGAGTTCA
TTTCTTTAGTGGGCATCCGTGACTCTCTGGACTCTGACCCATCTAACGCCTATCTG
TATTTACAAATACATTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT
ATCTATCTATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGGCTGCCTATGG
CTCAACCCAAGTTGAAGGAGGAGATTTGACCAACAATTCAAGCTCTCTGAATATGT
TTTGAA

(B) D18S51 (GenBank: AP001534.2)

GTTGAAAGATGAAATAACTTACTGAAATTGTTAATGAAGTATTGGATAAGCTACTT
TAAAAATAACAAACCCGACTACCAGCAACAACACAAATAAACAAACCGTCAGCCTA
AGGTGGACATGTTGGCTTCTCTCTGTTCTTAACATGTTAAAATTAAAATTAACTTC
TCTGGTGTGTGGAGATGTCTTACAATAACAGTTGCTACTATTTCTTTTCTTTTTCT
CTTTCTTTCCTCTCTCTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT
TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTGAGACAAGGTCTCAATTTGTC
ACTCAGAGTGAAGTGCAGTGGCATGAACATGGCTCACTGCAGCCTTAACCTTCTGG
GCTCAAGAACTCCTCCTGCCTCAGCCCTGCAAGTAGCTGAGACTACAGGCACGTGC

(C) D21S11 (GenBank: AP000433.2)

CTTGGACAGCCACACTGCCAGCTTCCCTGATTCTTCAGCTTGTAGATGGTCTGTTA
TGGGACTTTTCTCAGTCTCCATAAATATGTGAGTCAATTCCCCAAGTGAATTGCCT
TCTATCTATCTATCTATCTGTCTGTCTGTCTGTCTGTCTGTCTATCTATCTATATC
TATCTATCTATCATCTATCTATCCATATCTATCTATCTATCTATCTATCTATCTAT
CTATCTATCTATCTATCGTCTATCTATCCAGTCTATCTACCTCCTATTAGTCTGTC
TCTGGAGAACATTGACTAATACAACATCTTTAATATATCACAGTTTAATTTCAAGT
TATATCATACCACTTCATACATTATATAAAACCTTACAGTGTTTCTCCCTTCTCAG

Fig. 9.5. PCR primers for D13S631, D18S51, D21S11 STR loci.

Loci:	D13S631, D18S51, D21S11	
Primer mix:	Make the following primer mixes:	
D13S317-F	6 FAM 5'-5'-ACAGAAGTCTGGGATGTGGA	10 pmol/µl
D13S317-R	5'-GCCCAAAAAGACAGACAGAA	10 pmol/µl
D18S51-F	6 FAM 5'-CAAACCCGACTACCAGCAAC	10 pmol/µl

D18S51-R	5'-GAGCCATGTTCATGCCACTG	10 pmol/µl
D21S11-F	6 FAM 5'-TGAGTCAATTCCCCAAG TGAA	10 pmol/µl
D21S11-R	5'-AGTCAATGTTCTCCAGAGACAGAC	10 pmol/µl
Amgl-F	6 FAM 5'-CTGATGGTTGGCCTCAAGCCT	10 pmol/µl
Amgl-R	5'- ATGAGGAAACCAGGGTTCCA	10 pmol/µl

For each sample label four tubes A, B, C and D (one each for the STR locus and Amgl).

Reactions conditions:

PCR mix:	23 µl
Respective primer mix:	1 µl
Taq polymerase:	0.1 µl (0.5 units)
DNA:	1 µl (20ng)

Thermal cycling:

Initial denaturation:	1 minute at 94^{o}C
No. of cycles:	25
Denaturation:	48 seconds at 93^{o}C
Annealing:	48 seconds at 60^{o}C
Extension:	1 minute at 72^{o}C
Final extension:	3 minute at 72^{o}C

Pool 5 µl amplified product from each of the 4 reactions and use 1.3 µl of the pool.

HiDi formamide:	10 µl/sample
GeneScan Liz:	0.5 µl/sample
Denature in thermal cycler:	94^{o}C for 4 minutes

Run on a Genetic analyzer as per manufacturers instructions.

Some technical problems with STR analysis

PCR stutter products

In amplification of STRs the elongating DNA strand can slip one or more steps forward resulting in formation of end products that are one or more repeat units smaller in length. This is also called slippage synthesis. On gel electrophoresis the shorter amplified products appears as shadow bands or peaks often called stutter bands or peaks. The stutter products larger in length than the actual product can also result by backward slippage but these are uncommon.

Stutter bands/peaks are most prominent in the amplification of di-nucleotide repeats but these are not uncommon in PCR of tetra-nucleotide repeats (Fig. 12.6).

Twin peaks of "+A" and "–A" products

Taq polymerase has the property of adding +A at the 3' end of the elongating strand that results in formation of products one base pair longer. These products appear as twin peaks on a genetic analyzer.

Quantity of DNA

Too much or too little DNA in PCR can produce poor results. The quantification of DNA is discussed in chapter 2. The quantity of DNA is very critical for analysis on a genetic analyzer. The analyzer works best with DNA quantity ranging from 0.5-1ng per reaction.

The gel electrophoresis and silver staining are more robust as far as the quantity of DNA is concerned. The usual range for good results is 50-200ng per reaction. The minimum amount of DNA that can be seen on a gel by this method is ~20ng (Fig. 9.6).

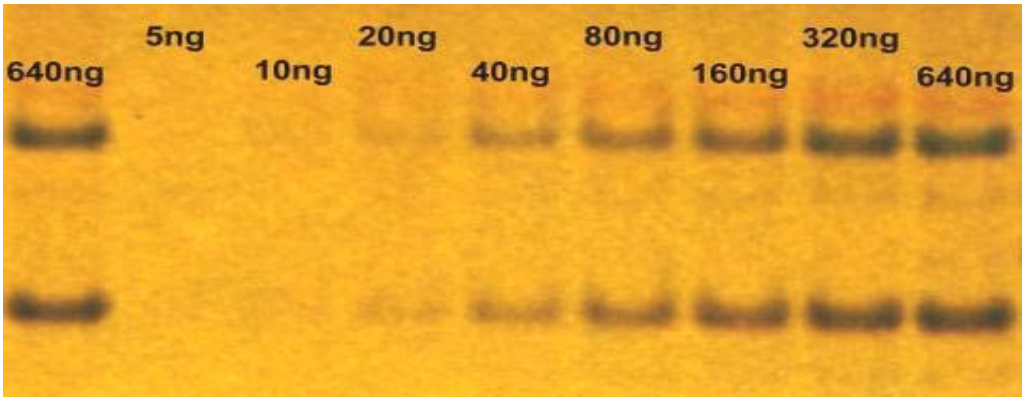

Fig. 9.6 Amplification of STR locus D8S1179 by using decreasing concentration of DNA. The 6% polyacrylamide gel was stained in silver nitrate. Minimum detectable amount of DNA is ~20ng.

Bibliography

1. Short Tandem Repeat DNA Internet DataBase, http://www.cstl.nist.gov/strbase/
2. Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 1998 62: 1408-1415.
3. Smith RN (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. *Biotechniques* 18: 122-128.