GENOMIC SEQUENCING OF ALLELES FROM STR LOCI.

L. Perlee and I. Balazs

Lifecodes Corporation, Stamford, CT, USA

Introduction.

The traditional procedures for the analysis of VNTR, AMP-FLP and STR loci, for DNA typing rely on the determination of the relative DNA fragment size for the definition of alleles. This type of allele definition provides large amount of information for VNTR loci containing substantial number of alleles. For loci composed of a few small alleles (i.e., a few hundred bases long), that are most suited for analysis by DNA amplification, the power of identity of a locus could be greatly expanded if additional source of variation were found at the level of their primary sequence. This would allow for further sub-classification of alleles for this type of loci. To this objective, we began studies on two tetranucleotide STR loci by sequencing their individual alleles.

Materials and Methods.

Genomic DNA from unrelated North American Caucasoid individuals was isolated from blood using proteinase K/SDS lysis and phenol extraction. Amplification was performed in a 100 μl reaction containing approximately 10 ng of DNA as template plus 200 μM dNTPs, 1X PCR buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$), 2.5 units of Taq polymerase (Perkin Elmer, Norwalk CT.) and 1 μM of primers. Cycling parameters were: 95 °C for 1', 50 °C for 1', and 72 °C for 1' for 30 cycles, followed by a 10 ' final extension at 72 °C. The amplified products were size fractionated in 3% high sieving agarose gel (, FMC Bioproducts) with 1 X Tris-Borate-EDTA buffer at 150 Volts for 5 hours.

The two STR selected for this study contain tetranucleotide repeats of the sequence GATA. The alleles detected contain from 8 to 16 repeats. Heterozygosity, determined from the frequency of different allele sizes was about 85 %.

After electrophoresis, individual alleles stained with ethidium bromide were excised from the gel, melted, diluted with H_2O and 10% used as template for a second amplification reaction. This second amplification was performed in two reactions using 5'biotinylated primers from opposite ends of the allele. The conditions used were the same used previously. The biotinylated amplification product was immobilized onto Streptavidin coated magnetic beads (Dynal, Inc., Glen Cove, NY) and the non-biotinylated strand eluted off with NaOH. The single stranded DNA was sequenced, in an ABI Model 373A automated sequencer, using fluorescent dideoxy terminators (Taq Dye-Deoxy Terminator Cycle Sequencing Kit; ABI, Foster City, CA) according to manufacturers' recommendations.

Results and Discussion.

The general strategy for sequencing individual STR alleles involved a primary amplification of genomic DNA, isolation of individual alleles from the gel after electrophoresis, followed by reamplification of each allele with a set of biotinylated primers. Single stranded DNA was isolated and sequenced.

The general property of these loci were:

Locus 1. Heterozygosity: 85%. STR: (GATA)n.

Number of alleles by size (by electrophoresis): 8

Frequency of allele number 1 = 4.2% Size range of alleles: 134 to 162 bp.
2 = 10.4%
3 = 10.4%
4 = 27.0%
5 = 18.8%
6 = 12.5%
7 = 8.3%
8 = 8.3%

Comparison of the sequence of several alleles within the same size class detected the existence of several sequence variants. For example, each of four alleles containing 13 GATA repeats showed sequence variations in the region contiguous to the repeat either due to sequence changes or insertion/deletions of a few bases.

$(GATA)_{13}$GAACGAACAC.....
$(GATA)_{13}$GATGAACAC.....
$(GATA)_{13}$GAACAC.....

Sequence analysis among alleles from the most common classes (i.e. 4, 5, 6) revealed 2 or 3 sequence variants in each group.

Locus 2. Heterozygosity: 85%. STR: (GATA)n.

Number of alleles by size (by electrophoresis): 10

Frequency of allele number 1 = 2.5% Size range of alleles: 220 to 300 bp.
2 = 7.5 %
3 = 5.0%
4 = 2.5%
5 = 17.5%
6 = 30.0%
7 = 15.0%
8 = 12.5%
9 = 5.0%
10 = 2.5%

Only 6 alleles have been subjected to sequence analysis. The results show that one class of genotypic variant was the presence of the sequence GATTA in the middle of the GATA repeats. Other changes detected represented mostly single base sequence changes. Comparison of two alleles containing the same number of GATA repeats show differences at various places in the sequence. For example:

(GATA)21-1: CAGGTCTCCGA.....GATACATAGATA.....TATAAACT
(GATA)21-2: CAGGTCT- CGA.....GATAGATAGATA.....TATACACT

(GATA)10-1: CAGGTCTTGA..... GATAGATAGATA.....TACCAACT
(GATA)10-2: CAGGTCCTGA..... GATAGATAGATA.....TACAAACT

The results obtained for both loci indicate significant variation at the DNA sequence level for alleles containing equal number of GATA repeats. This allows for the identification of new alleles within the same size class. The effect of this increase in the number of alleles on the frequency calculations will require analysis of a large number of alleles, however, it will result in an increase of the power of identification of these loci.

It is possible that the frequency of mutations may depend on the sequence of STR. However, there has not been enough of this type of loci sequenced to date to draw conclusions.

A study similar to the one described in this paper was reported by Adams et al. (1993) using an Alu-associated STR locus. The results of that study also shows large number of sequence variants among alleles with the potential of an increase in the power of identification of the locus.

References.

Adams M, Urquhart A, Kimpton C and Gill P (1993) The human D11S554 locus: four distinct families of repeat pattern alleles at one locus. Human Molecular Genetics 2:1373-1376