Analysis of sequence variations in the alleles from three STR loci.

L. Perlee, J. Neuweiler, I. Balazs

Lifecodes Corp., 550 West Ave., Stamford Ct 06902, USA

INTRODUCTION

DNA polymorphic loci resulting from variations in the number of short tandem repeats (STR), especially tetranucleotide repeats, are finding increased application in human identification. The main advantage of these loci are the small size of their alleles and their ability to be detected by PCR based DNA amplification. We have developed a multiplex system composed of 3 loci. The alleles from each locus fall within discreet size ranges and do not overlap. Although in theory the alleles from these loci differ by changes in the number of tetranucleotide repeats, additional variants may exist due to sequence variations between similar size alleles. Sequence studies, by Adams et al. (1993), of the alleles of the locus D11S554 have shown the occurrence of sequence variants among the same size alleles. The objective of this study was to report our initial findings on the primary sequence of multiple alleles from the 3 polymorphic STR loci.

MATERIALS AND METHODS

The alleles analyzed belong to the loci D3S1744, D12S1090 and D18S849.

DNA samples, from individuals possessing the allele(s) to be sequenced, were amplified with the locus specific primers (Lifecodes Corp.) using the conditions described by Neuweiler et al. (1995). The amplified alleles were fractionated by electrophoresis in an agarose gel. A portion of the gel containing the allele to be sequenced was sampled using a Pasteur pipette and the DNA re-amplified using one of the primer pairs 5'-labeled with biotin. DNA sequence was generated in both orientations by amplification with biotin-labeled primer for the 5' and 3' ends of the DNA fragment. The amplified product was purified by absorption to Streptavidin coated magnetic beads (Dynal, Inc) and sequenced using the Taq Dye-deoxy Terminator Cycle Sequencing kit (Perkin-Elmer) in an ABI Model 373A automated sequencer.

RESULTS AND DISCUSSION

DNA sequence analysis of alleles from the loci D3S1744, D12S1090 and D18S849 reveals that the main source of the size polymorphisms is the result of variations in the number of GATA repeat units.

Locus D18S849, consisted of alleles having from 9 to 19 tetranucleotide repeats. Alleles containing 10 and 13 repeats were not detected in the 200 individuals examined to date. In all allele sizes, the third repeat was a TATA followed by GAATA instead of GATAGATA. To determine if we could detect sequence differences between alleles of the same size, we selected the most common allele for locus D18S849. Alleles of the same size, from 8 unrelated individuals were sequenced. The results did not reveal differences in DNA sequence, however, a larger number of allele samples, from different populations, may have to be examined to determine the possible level of sequence

variation at this locus,

Locus D3S1744 contained from 14 to 22 GATA repeats and was flanked by short dinucleotide repeats with the sequence TC and TA respectively. In addition, all the allele sizes contained the sequence GAT, instead of the third GATA repeat.

Locus D12S1090 was the locus with the largest number of alleles. The number of GATA repeats varied from 9 to 33 in the different size alleles. These repeats were flanked by stretches containing TA repeats. Some alleles contained changes in the sequence such as G to C in the second repeat of many alleles. Some alleles also had one or two base insertions of T or TA or deletions of GA in the region of the repeats. A summary of the sequence of the STR region of these alleles have been summarized in Table 1. The number of different size alleles identified by gel electrophoresis was more than 25. Because of the large number of alleles at this locus, the allelic ladder made for the analysis of this locus was prepared using only every other allele. From approximately 400 alleles examined, 3 alleles could be classified as rare variants. These alleles had a 1 or 2 base difference with the common allele with allele sizes of 251, 255, and 285 bases (Table 1). Sequence analysis of same size allele from a limited number of unrelated individuals has not yet revealed sequence differences. A larger survey, however, may reveal sequence differences. The internal sequence variations observed between different size alleles suggest that this marker may be useful in studies of population diversity.

REFERENCES

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

Neuweiler J, Perlee L, Venturini J, Balazs I (1995) Properties of an STR multiplex marker system suitable for paternity and forensic determinations. this proceeding.

Table 1. Sequence of the STR region for alleles from the D12S1090 locus

ALLELE SIZE						
306	TATATATATATATATA	GATAGATA 15x(GATA)	GAT	2x(GATA)	2x(GATT) GATTA 10x(GATA)	GATGTTATAGATATATATTATAT
302	TATATATATATATATA	GATA*C*ATA 14x(GATA)	GAT	2x(GATA)	2x(GATT) GATTA 10x(GATA)	GATGTTATAGATATATATATAAT
298	TATATATATTTATATA	GATA*C*ATA 13x(GATA)	GAT	3x(GATA)	2x(GATT) GATTA 9x(GATA)	GATGTTATAGATATATATT AT
294	TATATATATTTATATA	GATA*C*ATA 13x(GATA)	GAT	2x(GATA)	2x(GATT) GATTA 9x(GATA)	GATGTTATAGATATATATT AT
290	TATATATATTTATATA	GATA*C*ATA 12x(GATA)	GAT	3x(GATA)	GATT GATTA 9x(GATA)	GATGTTATAGATATATATA AT
288	TATATATATATATATA	GATAGATA 13x(GATA)	GATTA	5x(GATA)TA	2x(GATA) GATTA 3x(GATA)	GATGTTATAGATATATATATAAT
285	TATATATATATATATA	GATA*C*ATA 13x(GATA)	GATTA	10x(GATA)		GATGTTATAGATATATATATAAT
284	TATATATATATATATA	GATAGATA 11x(GATA)	GATTA	5x(GATA)TA	2x(GATA)GATCAG 3x(GATA)	GATGTTATAGATATATATATAAT
277	TATATATATATATATA	GATA*C*ATA 13x(GATA)	GATTA	8x(GATA)		GATGTTATAGATATATATATAAT
269	TATATATATTTATATA	GATA*C*ATA 12x(GATA)	GATTA	7x(GATA)		GATGTTATAGATATATATATAAT
261	TATATATATATATATA	GATAGATA 11x(GATA)	GATTA	6x(GATA)		GATGTTATAGATATATATATAAT
257	TATATATATATATATA	GATAGATA 10x(GATA)	GATTA	6x(GATA)		GATGTTATAGATATATATATTAT
255	TATATATATTTATATA	GATAGATA 10x(GATA)TA	GATTA	5x(GATA)		GATGTTATAGATATATATATTAT
253	TATATATATTTATATA	GATAGATA 10x(GATA)	GATTA	5x(GATA)		GATGTTATAGATATATATATTAT
251	TATATATATATATATA	GATAGATA 9x(GATA) TA 2x(GATA)	GATTA	3x(GATA)		GATGTTATAGATATATATATAAT
244	TATATATATTTATATA	GATAGATA 14x(GATA)				GATGTTATAGATATATATATTAT
236	TATATATATTTATATA	GATA*C*ATA 12x(GATA)				GATGTTATAGATATATATATAAT
228	TATATATATTTATATA	GATA*C*ATA 10x(GATA)				GATGTTATAGATATATATATTAT
220	TATATATATTTATATA	GATAGATA 8x(GATA)				GATGTTATAGATATATATATTAT
212	TATATATATTTATATA	GATAGATA 6x(GATA)				GATGNTATAGATATATATATTAT

Underlined numbers, represent alleles used for the allelic ladder.
Numbers in bold, represent common alleles not present in the allelic ladder.
Numbers in italic, represent rare allelic variants