J Forensic Sci, Jan. 2003, Vol. 48, No. 1
Paper ID JFS2002065_481
Available online at: www.astm.org

TECHNICAL NOTE

Erik T. Bieschke,[1] M.S.; Margaret M. Wallace,[2] Ph.D.; Peter R. De Forest,[2] D. Crim.; Robert C. Shaler,[1] Ph.D.; and Mechthild Prinz,[1] Ph.D.

Characterization of a Novel Dimorphism in the 5′ Flanking Region of the Short Tandem Repeat (STR) Locus, *c-fes/fps* (FES)

ABSTRACT: The FES short tandem repeat (STR) locus contains seven to 14 repeats of the tetranucleotide sequence ATTT. A novel 10 base pair dimorphism in the 5′ flanking region of the FES locus was characterized in four broad populations: African-American, Hispanic, Caucasian, and Asian. The absence of the 10 base pair sequence, or (−) allele, was closely linked to FES STR alleles with 10 or fewer repeats. The presence of the 10 base pair sequence, or (+) allele, was closely linked to FES STR alleles with 12 or more repeats. The (−) and (+) alleles occurred equally often in FES STR allele11. The nucleotide sequence (5′-GGCTGTTTTG-3′) of the (+) allele, located 179 base pairs upstream of the FES STR, was determined to be consistent within and among the four populations. Statistical and sequence analysis confirmed the linkage between the two polymorphic sites. The results indicate that the exclusion rate of the FES locus is increased, above that for the STR alone, when both polymorphic characteristics are considered.

KEYWORDS: forensic science, DNA typing, FES, short tandem repeats, upstream flanking region, variants, alleles

Satellite DNA consists of tandemly repeated, non-coding sequences interspersed throughout the human genome. These polymorphic loci (1) form the basis of DNA typing systems currently used in human identification (2).

One class of satellite DNA, the microsatellites or short tandem repeats (STRs), consists of repetitive units containing two to five base pairs (3–5). The nature of the repeat region can be described as simple, compound, or complex. A simple repeat, such as the *c-fes/fps* proto-oncogene (FES) contains only one repeating sequence. The tetrameric FES STR has the repeat structure $(ATTT)_n$ and is located in intron 5 of the human *c-fes/fps* proto-oncogene on chromosome 15q25-qter (6).

Additional length and sequence polymorphisms occur in the repeat region, as well as in the supposedly constant flanking regions, within some of these loci. Subtypes in the repeat region have been reported at several loci: von Willebrand factor (VWA) (7) and tyrosine hydroxylase (THO1) (8). These variants involve alterations (base substitutions or deletions) within one or more of the repetitive units in the locus.

Length and sequence polymorphisms in the flanking regions have also been observed at the VWA, FES (7), and factor XIII (F13A1) loci (8–9). An A to C transversion has been described in the 5′ flanking region of the FES locus. This mutation is closely linked to alleles 10 and 11 (7). An additional subtype, the presence (+) or absence (−) of a 10 bp sequence, in the 5′ flanking region of the FES locus has been discovered by the present authors and is reported herein.

Materials and Methods

DNA was extracted from bloodstains collected at autopsy by the NYC Office of the Chief Medical Examiner using the Chelex method (10). Ethnic origin was determined by declaration by next of kin and/or skin color. The use of these human samples for research purposes was approved by the John Jay College of Criminal Justice Institutional Review Board.

Human DNA was quantitated using the method of Walsh et al. (11). DNA extracts were amplified using the GeneAMP® PCR System 9600 (PE Applied Biosystems, Division of Perkin Elmer, Foster City, CA) following the protocol of Klimpton et al. (12): 95°C for 11 min; 29 cycles of 95°C for 1 min, 54°C for 3 min, 72°C for 1 min; then 72°C for 10 min. The FES locus was amplified using two sets of custom-synthesized primers (Perkin Elmer): 1) Forward (FES/1): 5′-GGGATTTCCCTATGGATTGG-3′; Reverse (FES/2 FAM labeled): 5′-GCGAAAGAATGAGACTACAT-3′ (6); 2) Forward (FES/3): 5′-TCCAAGCTTTTGGCCACTAG-3′ (present study); Reverse (FES/2 FAM labeled). These primer pairs amplify nucleotide regions 4558–4780 and 4481–4780, respectively, within the human *c-fes/fps* locus (GenBank accession number X06292).

The FES/1-FES/2 primer pair (6) is a component of the quadruplex (QUAD) system developed by the British Forensic Science Service (12). The FES/3 primer was designed using the Primer Ex-

[1] Department of Forensic Biology, Office of the Chief Medical Examiner, New York, NY.
[1] Science Department, John Jay College of Criminal Justice, City University of New York, 445 West 59th Street, New York, NY.

Received 15 Feb. 2002; and in revised form 22 May and 30 June 2002; accepted 30 June 2002; published 13 Nov. 2002.

press software (Perkin Elmer). The outside primer pair, FES/3-FES/2, was used to amplify the upstream region which contained the dimorphic site. Using this primer pair, the resultant amplification products would be 77 and 87-bp longer that those obtained with the QUAD primers for the (−) and (+) alleles, respectively.

PCR products were electrophoresed using an ABI Prism 377™ DNA Sequencer (Perkin Elmer) according to the manufacturer's instructions. Results were analyzed using the GeneScan™, Version 2.1 and GenoTyper™, Version 2.0 software packages (Perkin Elmer). Size data from the samples amplified from the FES/3-FES/2 primer pair were compared to the size data from the same samples amplified using the FES/1- FES/2 (6,12) primer pair. FES alleles were designated according to the STR repeat number and the presence (+)/absence (−) of the upstream sequence. For example, an individual who was previously typed as an 11,11 homozygote could possibly be typed as an 11(−), 11(+) heterozygote. Herein lies the potential for increased discrimination at the locus.

FES allele data sets were compiled for four broad populations: African-American (N=99), Asian (N=105), Hispanic (N=104), and Caucasian (N=103). Statistical evaluations were performed on the data using the software program DNA-View, Version 22.09 (13). The exact test, an alternative to the χ^2 test, was used to check for Hardy-Weinberg equilibrium. Linkage equilibrium between the STR and dimorphic loci was tested using the exact test for independence. The exclusion rate, defined as the probability that a random sample will be excluded from a known type, was also calculated for each data set.

Sixteen samples, four from each ethnic group, were selected for sequencing. These samples were homozygous for both the STR and the dimorphism. Thus, the region could be sequenced from both directions, using the unlabeled primers FES/3 (forward) and FES/2 (reverse), without prior allele separation and purification. Amplified DNA was concentrated by centrifugal filtration using Micron-100 columns (Amicon, Division of Millipore, Bedford, MA) according to the manufacturer's instructions. The ABI Prism® dRhodamine Terminator Cycle Sequencing Kit (Perkin Elmer) was used for cycle sequencing following the manufacturer's instructions. Briefly, 20 μL reaction volumes containing 8 μL Terminator Ready Reaction Mix, 4.0 pmol forward or reverse primer, 30 to 90 ng of purified DNA were amplified [25 cycles of 96°C for 10 s, 56°C (forward primer) or 50°C (reverse primer) for 5 s, 60°C for 4 min] using the GeneAMP® PCR System 9600 (Perkin Elmer). Extension products were purified using the Centri-Sep (Princeton Separations, Adelphia, NJ) spin columns according to the manufacturer's directions.

The reaction products were electrophoresed using the ABI Prism™ 377 DNA Sequencer. Sequences were determined using the DNA Sequencing Analysis Software, Version 3.3 (Perkin Elmer). Alignments were constructed and compared using the subprograms Sequence Navigator, Version 1.0.1 and Factura, Version 1.2.0r6.

Results and Discussion

Amplification with the outside primers (FES/3-FES/2) yielded PCR products 77 or 87-bp larger than the product from the original primers (6,12). Each STR allele could be linked to one of two possible flanking sequences, thus, expanding the number of observed alleles from eight to 16 (Table 1).

As indicated in Table 1, FES STR allele 11 is the most common allele in all populations. The African-American and Hispanic populations have similar distributions. Within these populations the (−) and (+) alleles are almost evenly distributed (54%:46% and 56%:44%, respectively).

TABLE 1—*Allele frequencies and population statistics for the FES locus.*

Allele	African-American Allele Frequency (%)	Hispanic Allele Frequency (%)	Caucasian Allele Frequency (%)	Asian Allele Frequency (%)
7(−)	0.0	0.0	0.0	0.0
7(+)	1.0	0.0	0.0	0.0
8(−)	13.6	4.8	0.0	0.0
8(+)	0.0	1.0	0.5	0.0
9(−)	0.0	1.9	0.5	0.0
9(+)	4.6	0.5	0.0	0.5
10(−)	13.6	19.7	30.6	2.4
10(+)	4.0	1.9	1.0	1.9
11(−)	22.7	23.6	30.1	10.0
11(+)	12.6	19.2	11.7	28.1
12(−)	3.5	5.3	7.3	8.1
12(+)	16.2	18.8	14.1	25.7
13(−)	0.5	0.5	0.0	0.0
13(+)	7.1	2.9	3.9	21.0
14(−)	0.0	0.0	0.0	0.0
14(+)	0.5	0.0	0.5	2.4
All(−)	54.0	55.8	68.5	20.5
All(+)	46.0	44.2	31.5	79.5
Population Statistics	**African-American N = 99**	**Hispanic N = 104**	**Caucasian N = 103**	**Asian N = 105**
p-value				
HW di	0.522	0.226	0.944	0.671
LE STR + di	0.001	0.001	0.001	0.430
Exclusion rate				
STR	0.920	0.868	0.836	0.849
STR + di	0.965	0.947	0.917	0.927

HW: Hardy-Weinberg Equilibrium; LE: Linkage Equilibrium; di: dimorphism.

The Caucasian and Asian populations show distinct biases. The Caucasian population, as shown in Table 1, is biased such that 73% of all STR alleles are either 10 or 11, and 68% of all flanking region alleles are (−). The bias in the Asian population is such that 72% of all STR alleles are either 11 or 12, and 80% of all flanking region alleles are (+).

Statistical evaluations were performed on the data using the software program DNA-View, Version 22.09 (13). The *p*-value (exact test) for each data set (ethnic group) was determined. All the values are within the acceptable (0.05 to 0.95) range. Hence, the locus is in Hardy-Weinberg equilibrium. The exclusion rates for each data set were also calculated by the software program. As shown in Table 1, the exclusion rate increases for all populations, when considering both the STR and the flanking region, as opposed to the STR alone.

As shown in Table 1, the exact test for independence between FES STR alleles and the flanking region dimorphism yielded the following p-values for the tested ethnic groups: African-American (0.001), Caucasian (0.001), Hispanic (0.001), and Asian (0.430).

The data from the African-American, Caucasian, and Hispanic ethnic groups (i.e., $p < 0.05$) indicate that there is a deviation from the expected distribution of alleles when working from the assumption that the two polymorphic sites were not linked (linkage equilibrium). The low *p*-values in these three ethnic groups was

due to the fact that the STR 12 allele showed increased odds against being present in combination with so many (+) alleles. Also, the STR 10 allele showed increased odds against being present in combination with so many (−) alleles (data not shown).

In the Asian population, however, there were two distinct biases present as described above. Therefore, the observed distribution of high number STR alleles (i.e., 11–13) in combination with (+) alleles was consistent with the expectation of linkage equilibrium.

The statistical analysis for the African-American, Caucasian, and Hispanic populations verifies the linkage between the two polymorphic sites. Although the statistical analysis does not confirm the linkage between the two sites in the Asian population, this occurrence can be explained by the bias in allele frequencies at both sites in the population. However, direct sequencing indicates that the two sites are very close to one another on the same chromosome (data not shown).

Sequencing data were obtained from 16 individuals. The sequences for all (+)/(+) individuals were consistent, as well as those for all (−)/(−) individuals. The nucleotide sequence (5′-GGCTGTTTTG-3′) of the (+) allele occurs 179-bp upstream of the FES STR locus. Both polymorphic regions are located in intron 5 of the human *c-fes/fps* proto-oncogene (GenBank accession number X06292).

The existence of an A to C transversion in the 5′ flanking region of the FES locus has been previously reported (7). The reported data indicate that C is present in eight of 18 individuals with the 10 and 11 STR alleles. In this study, the polymorphic C was found in three of 11 individuals (data not shown).

The existence of polymorphic subtypes has forensic significance for three reasons. First, if the variant is readily detectable and amenable to routine analysis, then testing for this polymorphic trait can potentially increase the exclusion rate of that locus. Second, the presence of a subtype mutation in the primer annealing region can cause non-amplification of alleles and problems with interpretation. Several null alleles due to mis-pairing of the primer have been encountered in forensic samples (14–15). Third, hidden exclusions may occur, when individuals have the same apparent type, but differ with respect to the subtype.

Our results underscore the importance of primer design. Polymorphism in the flanking region can cause non-amplification of some alleles due to mis-pairing of the primer. If the FES primers were designed to span this 10 bp dimorphic region, false allelic profiles would be obtained for many individuals. Allelic and locus dropout would be prevalent. Laboratories developing forensic PCR systems should examine several alternative primer sites and combinations to verify the reproducibility of allelic profiles. Also, many diverse ethnic populations should be sampled in order to verify the consistency of the flanking region and absence of primer site mutations among these populations. If primer site mutations cannot be avoided, then the problem can be overcome by having degenerate primers in the reaction mix (14–15).

References

1. Wyman AR, White R. A highly polymorphic locus in human DNA. PNAS (USA) 1980;77:6754–675.
2. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
3. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the Polymerase Chain Reaction. Am J Hum Genet 1989;44:388–96.
4. Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 1989;44:397–401.
5. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.
6. Polymeropoulos MH, Rath DS, Xiao H, Merril CR. Tetranucleotide repeat polymorphism at the human fes/fps proto-oncogene (FES). Nucleic Acids Res 1991;19:4018.
7. Moller A, Meyer E, Brinkmann B. Different types of structural variation in STRs: HumFES/FPS, HumVWA, and HumD21S11. Int J Leg Med 1994;106:319–23.
8. Urquhart A, Kimpton CP, Downes TJ, Gill P. Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. Int J Leg Med 1994;107:13–20.
9. Puers C, Hammond HA, Caskey CT, Lins AM, Sprecher CJ, Brinkmann B, et al. Allelic ladder characterization of the short tandem repeat polymorphism located in the 5′ flanking region to the human coagulation factor XIII A subunit gene. Genomics 1994;23:260–4.
10. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR based typing from forensic material. BioTechniques 1991;10:506–13.
11. Walsh, PS, Valaro J, Reynolds R. A rapid chemiluminescent method for quantitation of human DNA. Nucleic Acids Research 1992;20:5061–5.
12. Kimpton C, Gill P, Walton A, Urquhart A, Millican ES, Adams M. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Meth Appl 1993;3:13–22.
13. Brenner CH. DNA-View: DNA Forensic Analysis Programs User's Manual. Version 21, Berkeley, CA, 1998.
14. Kline MC, Jenkins B. Non-amplification of a vWA allele. J Forensic Sci 1998;43(1):250.
15. Walsh S. Non-amplification of a vWA allele. Commentary on Kline MC, Jenkins B, Rodgers S. J Forensic Sci 1998;43:1103–4.

Additional information and reprint requests:
Margaret M. Wallace, Ph.D.
Science Department
John Jay College of Criminal Justice
City University of New York
445 West 59th Street
New York, NY 10019