Int J Leg Med (1994) 106:319–323

International Journal of
Legal Medicine

Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11

A. Möller, E. Meyer, B. Brinkmann

Institute of Legal Medicine, Westfälische Wilhelms-Universität, Von-Esmarch-Strasse 86, D-48149 Münster, Germany

Received April 13, 1994 / Received in revised form April 14, 1994

Summary. Alleles of the STR systems HumFES/FPS, HumVWA and HumD21S11 were sequenced and analyzed. Sequence data revealed 3 different systems concerning the complexity of their sequence structure. HumFES/FPS belongs to the STR polymorphism with a simple repeat structure. Only 2 subtypes were found with a base substitution in the 5′-flanking region and no variation in the repeat region. In the STR system HumVWA the sequence structure of the repeat region is more complex, because 2 tetranucleotide units TCTA and TCTG were present. Additionally allele 14 revealed a completely different sequence structure leading to a different electrophoretic mobility. The repeat region of HumD21S11 is compound in structure. The possibility of variation at 3 positions leads to the occurrence of microheterogeneities in fragments of apparent length. In the upper allele range alleles arise with an additional incomplete TA-repeat.

Key words: STR systems HumFES/FPS/HumVWA/HumD21S11 – Sequencing – Sequence structure – Allelic ladder

Zusammenfassung. Allele der STR Systeme HumFES/FPS, HumVWA and HumD21S11 wurden sequenziert and analysiert. Die Sequenzierdaten zeigen 3 veschiedene Systeme die Komplexität ihrer Sequenzstrukturen betreffend. HumFES/FPS ist ein STR Polymorphismus mit einer einfachen Repeatstruktur. Nur 2 Subtypen mit einer Basensubstitution im 5′-flankierenden Bereich wurden gefunden und keine Sequenzvariation in der Repeatregion. Die Sequenzstruktur des STR Systems HumVWA ist komplexer aufgrund der 2 Tetranukleotideinheiten TCTA und TCTG. Zusätzlich zeigt das Allel 14 eine vollständig andere Sequenzstruktur, die zu einer unterschiedlichen elektrophoretischen Mobilität führt. Die Repeatregion von D21S11 besitzt eine komplizierte Struktur. Die Möglichkeit der Variation an 3 unterschiedlichen Positionen führt zu dem Auftreten von Mikroheterogenitäten in Allelen von gleicher Länge. Im oberen Allelbereich treten Allele mit einem zusätzlichen unvollständigen TA-Repeat auf.

Correspondence to: B. Brinkmann

Schlüsselwörter: STR Systeme HumFES/FPS/HumVWA/HumD21S11 – Sequenzierung – Sequenzstruktur – Allelische Leiter

Introduction

STR (short tandem repeat) systems are highly polymorphic marker systems which offer many advantages for DNA typing (Edwards et al. 1992; Brinkmann 1992; Kimpton et al. 1992; Wiegend et al. 1993; Möller et al. 1994). Apart from the regular repeat structure (consensus structure) there can exist different further types of either length or sequence variation, both in the repeat region and in the flanking region (Puers et al. 1993; Urquhart et al. 1993; Möller and Brinkmann 1994; Adams et al. 1993). Such variations can grossly influence the accuracy and precision of allele definition and also, the reproducibility (Möller and Brinkmann 1994). On the other hand, precise knowledge of such microheterogeneities is the prerequisite of the definition of allelic ladders for all systems. With this aim, we have investigated the sequence structure of the STRs HumFES/FPS (Polymeropoulos et al. 1991), HumVWA (Kimpton et al. 1992), HumD21S11 (Sharma and Litt 1992).

Materials and methods

DNA was extracted from blood samples of Caucasians from the Münster area as described (Brinkmann et al. 1991) and quantified using the slot blot technique (probe D17Z1, Gibco BRL, Waye et al. 1989).

PCR amplification:

HumVWA: as described by Möller et al. (1994)

HumFES/FPS: 5 ng template DNA, 1 U Taq DNA polymerase, 1 μM each primer, 200 μM each nucleotide, 2 ml 10 × PCR buffer (Promega Corporation, USA) diluted with double distilled water to a final volume of 25 µl and overlayed with 30 µl mineral oil.

HumD21S11: 2 ng template DNA, 1 U Taq DNA polymerase, 1 μM each primer, 100 μM each nucleotide, 2 µl 10 × PCR buffer

allele designation (1)	fragment length (bp) (2)	sequence structure	number of sequenced alleles
8	211	5'-FR * A (ATTT)$_8$ 3'-FR **	2
9	215	A (ATTT)$_9$	1
10a	219	C (ATTT)$_{10}$	3
10	219	A (ATTT)$_{10}$	5
11a	223	C (ATTT)$_{11}$	5
11	223	A (ATTT)$_{11}$	5
12	227	A (ATTT)$_{12}$	2
13	231	A (ATTT)$_{13}$	1

Fig. 1. Sequence structure of the STR system HumFES/FPS. Underlined alleles are included in the allelic ladder (Fig. 2). (1) allele designation according to the repeat number, the alleles 10a and 11a are named according to their faster electrophoretic mobilities to the anode (a = anodal); (2) fragment length in base pairs (bp) as determined by Taq-Cycle-Sequencing. * 5′-flanking region, 155 bp; ** 3′-flanking region, 24 bp; ▶ sequence variation at position 34 in the 5′-flanking region

(Promega Corporation, USA), 200 ng/µl BSA (bovine serum albumin); diluted with double distilled water to a final volume of 25 µl and overlayed with 30 µl mineral oil.

Primer sequences:

HumVWA (Kimpton et al. 1992)
P1: 5′CCC TAG TGG ATG ATA AGA ATA ATC 3′
P2: 5′GGA CAG ATG ATA AAT ACA TAG GAT GGA TGG 3′

HumFES/FPS (Polymeropoulos et al. 1991)
P1: 5′GGG ATT TCC CTA TGG ATT GG 3′
P2: 5′GCG AAA GAA TGA GAC TAC AT 3′

HumD21S11 (Sharma and Litt 1992)
P1: 5′GTG AGT CAA TTC CCC AAG 3′
P2: 5′GTT GTA TTA GTC AAT GTT CTC 3′

Amplification conditions:

HumVWA: according to Möller et al. (1994)
HumFES/FPS: 95° C–1 min, 54° C–1 min, 72° C–1 min; 30 cycles with an additional extension step of 72° C for 10 min (Biometra Triothermoblock, Germany)

HumD21S11: 94° C–15s, 62° C–30s, 72° C–75s; 29 cycles (Biometra Triothermoblock, Germany)

Non-denaturing gels and electrophoresis: Polyacrylamide (6% T, 3% C, 750 µm), piperazine diacrylamide as crosslinker, 28 m*M* CHES (Cyclohexylaminoethane sulfonic acid), 81 m*M* formate (pH 9.0). – 18 cm separation distance (HumVWA, HumFES/FPS), 16 cm separation distance (HumD21S11), 2% agarose plugs in 2 × tris (0.5 *M*)/borate buffer. – Initial 1000 V, 40 mA, 5 W with ramping every 90 min up to 15 W which was continued until the bromophenol blue marker had reached the anode. Bands were visualized by silver staining (Budowle et al. 1991).

Denaturing gels and electrophoresis: The alleles of HumFES/FPS and HumVWA were also separated on a 6% denaturing gel (8.3 *M* urea, 6% acrylamide/bis-solution and 1 × tris/borate buffer, separation distance 12 cm, 800 V, 45 m*M*, 30 W) using the fluorescence detection system (Applied Biosystems, Foster City, CA). The fragment sizes were automatically analyzed using the internal standard Genescan 2500 labelled with ROX (6-carboxyrhodamin X) and the Genescan software 672 (Applied Biosystems, Foster City, CA).

Taq-Cycle-Sequencing and analysis of sequence data:
Isolation of DNA fragments from the gel after silver staining, Taq-Cycle-Sequencing and analysis of the sequence data was performed as previously described (Möller and Brinkmann 1994). The PCR primers were used as sequencing primers.

Allele designation: Allele designation of the STR systems HumVWA and HumFES/FPS was according to the repeat number

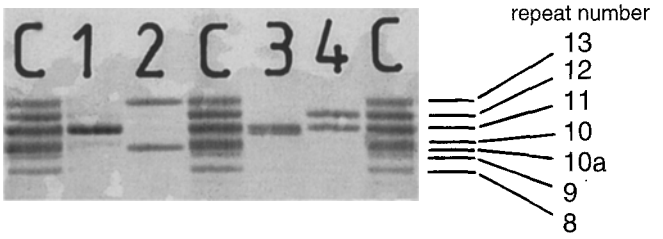

Fig. 2. HumFES/FPS. Allelic ladder composed of 7 sequenced alleles. Samples were separated on a high-resolution/non-denaturing gel and visualized by silver staining (Budowle et al. 1991). C = allelic ladder (designation according to Fig. 1; 1 = allele No. 11; 2 = alleles No. 10a, 13; 3 = alleles No. 11a, 11; 4 = alleles No. 11, 12)

(ISFH Recommendations 1992; ISFH Recommendations 1993). For HumD21S11, which has a complex sequence structure, alleles were designated according to the total number of tetranucleotide units in the repeat region consisting of TCTR (Nomenclature committee of the International Union of Biochemistry 1985). According to the recommendations interalleles have been designated by a suffix, indicating the number of bases of the incomplete repeat.

Results

HumFES/FPS

A total of 25 alleles were sequenced and analyzed (Fig. 1). All alleles showed a regular repeat structure with the number of repeats varying between 8 and 13. The 5′-flanking region (FR, 155 bp) showed a transversion from A to C at position 34 in the alleles 10 and 11, while in the 3′-flanking region (24 bp) no variation was found. The variant alleles called 10a and 11a occurred with a frequency of 3 out of 8 (10a) and 5 out of 10 (11a) respectively. The electrophoretic mobilities of the variant alleles 10a/11a differed from the regular ones (10/11) in high-resolution/non-denaturing gels (Fig. 2), while they were indistinguishable in a denaturing gel system.

HumVWA

A total of 50 alleles were sequenced and analyzed (Table 1). The flanking regions (FR) consisted out of 43 bp (5′-FR) and 39 bp (3′-FR) respectively. All alleles sequenced so far showed the same basic repeat structure TCTA

Table 1. Sequence structure of the repeat region, fragment lengths and allele designation of the STR system HumVWA

Allele designation (1)	Fragment length (bp) (2)	Sequence structure	Number of sequenced alleles
13	134	TCTA $(TCTG)_4$ $(TCTA)_8$	2
14	138	TCTA TCTG $(TCTG)_4$ $(TCTA)_3$ TCCA $(TCTA)_3$	12
15	142	TCTA $(TCTG)_4$ $(TCTA)_{10}$	5
15′	142	TCTA $(TCTG)_3$ $(TCTA)_{11}$	1
16	146	TCTA $(TCTG)_4$ $(TCTA)_{11}$	5
16′	146	TCTA $(TCTG)_3$ $(TCTA)_{12}$	1
17	150	TCTA $(TCTG)_4$ $(TCTA)_{12}$	6
18	154	TCTA $(TCTG)_4$ $(TCTA)_{13}$	6
19	158	TCTA $(TCTG)_4$ $(TCTA)_{14}$	5
20	162	TCTA $(TCTG)_4$ $(TCTA)_{15}$	4
21	166	TCTA $(TCTG)_4$ $(TCTA)_{16}$	2
22	170	TCTA $(TCTG)_4$ $(TCTA)_{17}$	1

Underlined alleles are included in the allelic ladder. (1) allele designation according to the repeat number; (2) fragment length in base pairs (bp) as determined by Taq-Cycle-Sequencing

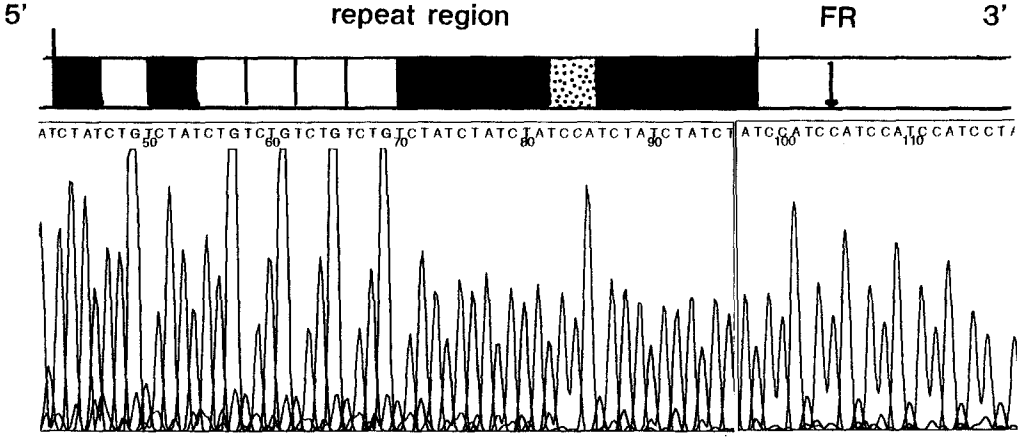

Fig. 3. Electropheretogram of the non-consensus allele No. 14 at the locus HumVWA in the white Caucasian population. Sequence data were obtained after separation of the samples in a 6% denaturing gel and automatically analyzed on the ABI 373A Sequencer. (■) TCTA, (▩) TCCA, (□) TCTG, (—) part of the 3′-flanking region with the base substitution at position 7 (arrow). FR = flanking region

$(TCTG)_{3-4}$ $(TCTA)_n$ except allele 14 which revealed a strongly different sequence structure leading to a different electrophoretic mobility in a high-resolution/non-denaturing gel system (Fig. 4). Differences occurred at 3 sites (Fig. 3; Table 1): (1) Before the beginning of the basic repeat structure (read from 5′) 2 tetranucleotides are additionally inserted, (2) the 4th repeat of the $(TCTA)_n$ shows a T to C transition, (3) the 3′-FR shows a T to C transition at position 7. – So far, 12 alleles 14 (from 12 unrelated individuals) have been sequenced, all of them showing the same structure. However none of the other alleles exhibited any of these alterations. – In addition, sequence variation was found in 2 alleles (15′, 16′, see Table 1) which were indistinguishable from the consensus variants (15 and 16) in high resolution/non-denaturing and denaturing gels.

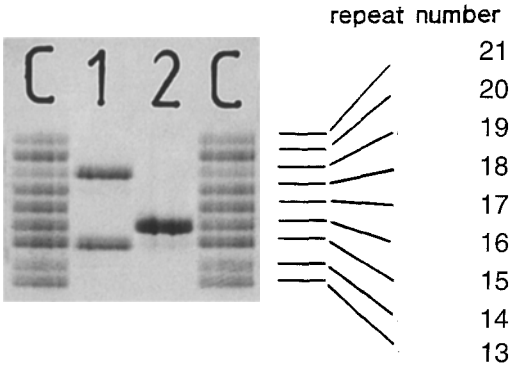

Fig. 4. HumVWA. Allelic ladder composed of 9 sequenced alleles. Samples were separated on a high-resolution/non-denaturing gel and visualized by silver staining. C = allelic ladder (designation according to Table 1; 1 = alleles No.15, 19; 2 = allele No. 16)

HumD21S11

A total of 38 alleles were sequenced and analyzed (Table 2). Fragment lengths ranged from 209 to 243 bp. The 5′-FR consisted out of 31 bp, the 3′-FR was 63 bp. Within the repeat region, 3 variable regions (vr = I, II, III, see Table 2) and 1 constant region (cr) located between II and III could be distinguished. Furthermore, part of the upper allele range (allele > 29) showed an incomplete repeat, ei-

Table 2. Repeat structure, fragment lengths and allele designation of the STR system HumD21S11

Allele designation (1)	Fragment length (bp) (2)	Sequence structure I	II	III	Number of sequenced alleles
26	209	$(TCTA)_4$	$(TCTG)_6$ **********	$(TCTA)_8$	5
27	213	$(TCTA)_4$	$(TCTG)_6$ **********	$(TCTA)_9$	2
28	217	$(TCTA)_4$	$(TCTG)_6$ **********	$(TCTA)_{10}$	3
29	221	$(TCTA)_4$	$(TCTG)_6$ **********	$(TCTA)_{11}$	1
29′	221	$(TCTA)_6$	$(TCTG)_5$ **********	$(TCTA)_{10}$	3
30	225	$(TCTA)_6$	$(TCTG)_5$ **********	$(TCTA)_{11}$	4
31′	229	$(TCTA)_6$	$(TCTG)_5$ **********	$(TCTA)_{12}$	3
31	229	$(TCTA)_5$	$(TCTG)_6$ **********	$(TCTA)_{12}$	1
32	233	$(TCTA)_6$	$(TCTG)_5$ **********	$(TCTA)_{13}$	1
30.2	227	$(TCTA)_5$	$(TCTG)_6$ **********	$(TCTA)_{10}$ TA TCTA	3
31.2	231	$(TCTA)_5$	$(TCTG)_6$ **********	$(TCTA)_{11}$ TA TCTA	4
32.2	235	$(TCTA)_5$	$(TCTG)_6$ **********	$(TCTA)_{12}$ TA TCTA	3
33.2	239	$(TCTA)_5$	$(TCTG)_6$ **********	$(TCTA)_{13}$ TA TCTA	4
34.2	243	$(TCTA)_5$	$(TCTG)_6$ **********	$(TCTA)_{14}$ TA TCTA	1
				$(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCA TA	

Underlined alleles are included in the allelic ladder. (1) allele designation, including the tetranucleotides TCTR in the repeat region; (2) fragment length in base pairs (bp) as determined by Taq-Cycle-Sequencing. cr = constant region (*****); I-III = variable region (vr)

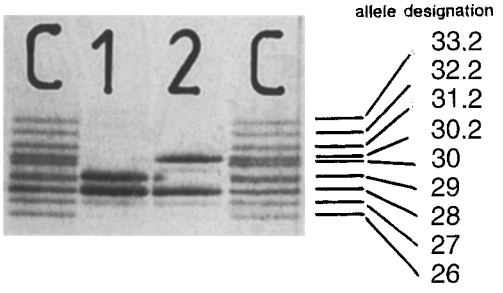

Fig. 5. HumD21S11. Allelic ladder composed of 9 sequenced alleles. Samples were separated on a high-resolution/non-denaturing gel and visualized by silver staining. C = allelic ladder (designation according to Table 2; 1 = alleles No. 28, 29; 2 = alleles No. 28, 30.2)

ther caused by a TA insertion or a TC deletion between the last 2 repeats at the 3′-end. The vr I consisted out of a $(TCTA)_{4-6}$ block, vr II out of a $TCTG_{5-6}$ block and the cr was a 43 bp array between II and III with hexa-, tri- and dinucleotides as well as further TCTA units. Vr III showed the highest variation in repeat number, thus representing the actual variable region. The incomplete repeat occurred in 5 subsequent alleles (30.2 to 34.2) and the site of insertion/deletion was the same in all alleles studied. The structural variants 29′ and 31′ (Table 2) could not be resolved from the consensus variants 29 and 31, neither in high-resolution/non-denaturing gels nor in denaturing gels. The alleles 30.2 to 34.2 could be resolved from their neighbouring consensus alleles (Fig. 5) in the gel system.

Discussion

Sequence data of the 3 STR systems HumFES/FPS, HumVWA and HumD21S11 were analyzed to generate allelic ladders of known sequence as required by the ISFH recommendations and to get more knowledge about possible microheterogeneities in these STR systems.

HumFES/FPS in Caucasians seems to be a STR polymorhpism with a constant basic structure. Only 2 structural variants were observed (10a, 11a) with slight but distinct migration differences in comparison to the consensus alleles only under non-denaturing conditions. From a forensic point of view it is important to know that the use of denaturing conditions could mask these variants: A visual match could therefore include the possiblitiy of a structural mismatch, but this possible exception seems to be restricted to 2 alleles. In non-denaturing gels the 2 alleles are distinguishable and therefore, we do not see any serious argument against the application of this system.

The HumVWA system shows a more complex repeat structure than HumFES/FPS with 2 tetranucleotide units TCTA and TCTG. The vast majority of alleles has the following sequence structure: TCTA $(TCTG)_4$ $(TCTA)_n$ with only very few exceptions: TCTA $(TCTG)_3$ $(TCTA)_n$. These variants could not be electrophoretically resolved from the consensus variants in the elctrophoretic system applied. As long as an allele-ladder-match exists the presence of a structural mismatch at one site can be disregarded. – In addition there exists a striking peculiarity, i.e. allele 14 which differs at 3 sites from the basic structure present in all other alleles investigated so far. Since none of these variations have yet been observed in any of the other alleles we tentatively assume that this allele has to be regarded as ancient and could have arisen by genetic drift. Concerning the evolution of such a non-consensus allele it would be of great interest to study its appearence in other human populations and e.g. in primate ancestors. Comparative analyses could possibly elucidate this problem. Furthermore, alleles 15′ and 16′ seem to be quite rare and

this type of variation might also occur in other allelic classes when more alleles have been investigated.

HumD21S11 is highly complex in structure and apart from the variations observed so far further variants might be detected. In the lower allelic range (alleles 26–29) there seems to be a regular basic structure with $(TCTA)_4$ $(TCTG)_6$ $(TCTR)_n$. Higher weight alleles contain insertions of either 1 or 2 tetranucleotide TCTA-units in vr I and the insertion of 2 such repeats is always associated with a deletion of 1 TCTG-repeat in vr II. Furthermore, from allele 30 upwards there occurrs an obvious branching into one subsequent series (up to 34.2) always showing an insertion of a dinucleotide at the same position and, another branch (up to 32) lacking this insertion. – Thus, this system is not only extremely polymorphic but also highly interesting for evolutionary biology studies. In forensic investigations a system is required which can resolve alleles but for practical reasons an allelic ladder should not contain all variants that exist. We propose the use of a consensus ladder such as that described here. In addition it has to be checked whether further variants can be resolved which subsequently have to be defined as well. In this case one could of course apply a match and bin approach which should be validated before the practical application.

Acknowledgements. The authors would like to thank A. Jürling and M. Schürenkamp for excellent technical assistance.

References

Adams M, Urquhart A, Kimpton C, Gill P (1993) The human D11S554 locus: four distinct families of repeat pattern alleles at one locus. Hum Mol Genet 2:1373–1376

Brinkmann B, Rand S, Wiegand P (1991) Population and family data of RFLP's using selected single- and multi-locus systems. Int J Leg Med 104:81–86

Brinkmann B (1992) The use of STRs in stain analysis. In: Proceedings from the Third International Symposium on Human Identification. Promega Corporation, Madison, USA, pp 357–373

Budowle B, Chakraborty R, Guisti AM, Eisenberg AJ, Allen RC (1991) Analysis of the variable number of tandem repeat locus D1S80 by the polymerase chain reaction followed by high resolution polyacrylamide gel electrophoresis. Am J Hum Genet 48:137–144

DNA recommendations-1992 report concerning recommendations of the DNA Commission of the International Society of Forensic Haemogenetics relating to the use of PCR-based polymorphism. Int J Leg Med 105:63–64

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12: 241–253

Kimpton CP, Walton A, Gill P (1992) A further tetranucleotide repeat polymorphism in the vWF gene. Hum Mol Genet 1:28

Möller A, Brinkmann B (1994) Locus ACTBP2 (SE33): Sequencing data reveal considerable polymorphism. Int J Leg Med 106:262–267

Möller A, Wiegand P, Grüschow C, Seuchter SA, Baur MP, Brinkmann B (1994) Population data and forensic efficiency values for the STR systems HumVWA, HumMBP and HumFABP. Int J Leg Med 106:183–189

Nomenclature Committee of the International Union of Biochemistry (NC-IUB) (1985) Nomenclature for incompletely specified bases in nucleic acid sequences. Eur J Biochem 150 (1–5): 4–8

Polymeropoulos MH, Rath DS, Xiao H, Merril RM (1991) Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). Nucleic Acids Res 19:4018

Puers C, Hammond HA, Jin L, Caskey T, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HumTH01 (AATG) and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–958

Sharma V, Litt M (1992) Tetranucleotide repeat polymorphism at the D21S11 locus. Hum Mol Genet 1:67

Urquhart A, Kimpton CP, Gill P (1993) Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Hum Genet 92:637–638

Waye JS, Lawrence PA, Budowle B, Shuttler GG, Fourney RM (1989) A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts. Biotechniques 7: 852–855

Wiegand P, Budowle B, Rand S, Brinkmann B (1993) Forensic validation of the STR systems SE33 and TC11. Int J Leg Med 105:315–320