Int J Legal Med (1997) 110 : 150–154

ORIGINAL ARTICLE

G. T. Duncan · K. Balamurugan · B. Budowle
J. Smerick · M. L. Tracey

Microvariation at the human D1S80 locus

Received: 16 August 1996 / Received in revised form: 17 December 1996

Abstract The minisatellite locus D1S80, (location: 1p35–p36), GenBank sequence accession # D28507), is a variable number of tandem repeat (VNTR) locus with a 16 base pair repeat size. The sequence of the predominant core repeat region and variants of the D1S80 locus were determined to ascertain whether sequence variation or size variation is the cause of altered migration of some D1S80 alleles. A total of 23 alleles from 14 individuals, previously typed based on the number of repeats (i.e. nominal alleles) for the D1S80 locus, were selected for sequence analysis. The individuals were from African American, Caucasian, and Hispanic databases. From these, 18 different repeat unit sequences were observed and arbitrarily designated A–R. Structural relationships between the alleles became more apparent when the arrays of repeat units were divided into common motifs or super-repeat domains. Six motifs ranging from 3 to 9 repeat units were identified. Several of the alleles included repeat arrays which were too diverse to predict an evolutionary relationship, however, there are two general repeat motif arrays and each has some relationship with either the 18 or the 24 repeat allele. The D1S80 allelic polymorphism is primarily due to variation in the number of repeat units and to sequence variation among repeats, however, it can not be ruled out that some rare alleles may be due to insertions or deletions.

Key words D1S80 locus · Interallele · Minisatellite sequencing · VNTR · Variant allele

G. T. Duncan (✉) · M. L. Tracey
Department of Biology, Florida International University, University Park Campus, Miami, Florida 33199, USA
FAX: +1 (954) 831 6413
e-mail: gdunca01@servax.fiu.edu

K. Balamurugan
Indianapolis-Marion County Forensic Services Agency, 40 South Alabama Street, Indianapolis, Indiana 46204, USA

B. Budowle · J. Smerick
Forensic Science Research and Training Center, FBI Academy, Quantico, Virginia 22135, USA

Introduction

The minisatellite locus D1S80, (1p35-p36), (GenBank sequence Accession # D28507), is a variable number of tandem repeat (VNTR) locus whose repeat size is 16 base pairs [1, 2]. With alleles defined by the number of repeat units, the D1S80 locus is highly polymorphic. The total number of alleles found in any one population sample has been reported as high as 28, however, the true variation is underestimated [3, 4].

Originally D1S80 was described as a marker with discrete alleles [2]; however, sequence variation and/or minor length variation within these discrete allelic classes has been suggested [5]. Some allelic bands were observed to migrate anodally, cathodally or medially with regard to common nominal alleles contained within an allelic marker ladder [6, 7]. These "off ladder" allelic variants, sometimes referred to as "interalleles" [8] were detectable generally under conditions where polyacrylamide gels are cooled (15° C) during electrophoretic separation of D1S80 amplicons [6, 9]. Various methods have been used to detect minor variants but sequencing remains the definitive method of choice to describe the variation at this locus.

A number of recent studies have centered on the sequence structure and variation that exist in STRs (short tandem repeat loci) including the ACTBP2 (SE33) [8, 10], the HumFES/FPS, HumVWA, and HumD21S11 loci [11, 12]. In addition the sequence structure of several minisatellite loci have also been characterized including CEB1(D2S90) [13], MS32 [14], MS205(D16S309) [15], and Apolipoprotein B 3′VNTR [16]. A common theme of these studies is the elucidation of common motifs consisting of groups of repeat units which appear in the sequence structure as sets of length alleles. It is the aim of this study to elucidate the structure of the D1S80 locus.

Reynolds (personal communication) determined the sequence of a number of common D1S80 alleles and two "off ladder" variants and found 15 distinct 16 bp repeat unit sequences. Despite two conserved blocks of repeat units, the alterations in the remaining repeat units ap-

peared to be random and variations in sequence as opposed to length variation caused variants to alter their migration compared with allelic ladder components.

The current study describes the sequence of the repeat region of the D1S80 locus and arrangement of the repeat units in allelic variants to determine whether or not sequence variation or size variation is the cause for altered migration of some D1S80 alleles.

Material and methods

Sample preparation

Whole blood samples from unrelated individuals were dried on to sterile cloth or S&S # 903 paper (Schleicher & Schuell, Keene, NH). Race/ethnicity of the subjects was determined by appearance and/or based on self-declaration classification. One gorilla blood sample was also included and was kindly supplied by Miami Metrozoo. The samples were stored at –20° C until analysis. DNA was extracted according to a previous method [17] and the quantity of recovered DNA was estimated using a slot blot human alphoid hybridization technique [18].

PCR amplification and sequencing of VNTR alleles

Amplification of the D1S80 locus was performed using the D1S80 AmpliFLP kit (Perkin Elmer, Norwalk, CT) or by a method previously described using a DNA Thermal cycler Model 9600 (Perkin Elmer, Norwalk, CT) [19]. Alleles were separated by discontinuous vertical and/or horizontal polyacrylamide gel electrophoresis. Allele designations were made by comparison with an allelic ladder run in adjacent lanes. From the first PCR 1 µl of amplicon was reamplified with oligonucleotide primers (forward, 5′ CGTACT*G-AATTC*GAAACTGGCCTCCAAACACTGCCCGCCG 3′ and reverse, 5′ CAGCTA*TCTAGA*GTCTTGTTGGAGATGCACGTGCCCCTTGC 3′) that contained the restriction sites *Eco*R I and *Xba* I, respectively *as underlined*. Amplicons were purified with Centricon 100 microconcentrators (Amicon, Danvers, MA), digested with *Eco*R I and *Xba* I and cloned in PBS(+) plasmid vector (Stratagene, La Jolla, Calif.). The XL1 blue strain of *E.coli* cells were transformed with the recombinant plasmid, and the colonies were screened for the presence of recombinants. The nucleotide sequence was determined by the dideoxy chain termination method [20] with the ABI Prism DNA sequencing kit using a 373A automated DNA sequencer (Applied Biosystems, Foster City, Calif.). Sequences were analyzed with the DNAstar software analysis package (Madison, Wisc.) for DNA analysis.

Results and discussion

Under our protocol, vertical gels attain temperatures during separation ranging from approximately 35–50° C, while horizontal gels are maintained at 15° C during electrophoretic separation. Off ladder variants were more prevalent when separated on horizontal gels. Most of these off ladder variants migrated to the same positions relative to the ladder alleles when typed using a vertical gel format.

There are three potential explanations for the observed polymorphism at the D1S80 locus i.e. 1) variation in the number of repeat sequences contained within the alleles, 2) nucleotide substitutions resulting in different sequences of some repeat units among individual alleles, and 3) insertions or deletions. The variation in the number of repeat units is what generally is typed at the D1S80 locus. However, the off ladder variants may more likely be the result of sequence variation and/or insertions or deletions. Because the microvariants were observed after electrophoretic separation in a lower temperature environment (i.e. potentially affecting secondary conformation of the microvariants), differences in sequence variation are the more plausible explanation.

To confirm this hypothesis, 23 selected alleles were sequenced. Different repeat units contained within an allele and repeat units between different alleles varied in their respective sequences and 18 different repeat unit sequences were observed and arbitrarily designated A–R (Table 1). Excluding the type A repeat unit, which is com-

Table 1 Nucleotide sequences of observed repeat units. Each repeat unit is assigned a letter code. (.) represents a match to the consensus sequence as represented by the Type H repeat unit. #OBS refers to the number of repeat units of this type observed in this study

Type	#OBS	Repeat unit																	
Type A	23	T	C	A	.	C	.	.	.	–	A	.	.	–	.	.	.		
Type B	23	A	C	A	.	.	.	.	.	.	A	.	.	.	.	.	.		
Type C	50	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.		
Type D	41	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.		
Type E	24	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.		
Type F	12	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.		
Type G	39	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.		
Type H	59	G	A	G	G	A	C	C	A	C	C	G	G	C	A	A	G		(Consensus)
Type I	136	.	.	.	.	.	.	.	.	.	.	A	.	G	.	.	.		
Type J	38	.	.	.	A	.	.	.	.	.	.	A	.	G	.	.	.		
Type K	9	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.		
Type L	23	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.		
Type M	1	.	.	.	.	.	.	.	.	.	.	A	.	G	.	G	.		
Type N	1	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.		
Type O	1	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.		
Type P	1	.	.	A	.	.	.	.	.	.	.	.	.	A	.	G	.		
Type Q	1	.	.	A	.	.	.	.	.	G	.	.	.	A	.	.	.		
Type R	1	.	G	.	A	.	.	.	.	.	.	A	.	G	.	.	.		

Table 2 Repeat unit array organization in the human D1S80 locus. The repeat length is directly related to the number of repeat units. The letter to the right of the number of repeat units is the position the allele falls anodally or cathodally to the nominal allele. The letter in parentheses identifies the broad population classification of the origin of the sample; C = Caucasian, H = Hispanic, and AA = Afro American. Each repeat unit is represented by a letter code (Table 1). B* represents a B repeat unit containing a deletion of a guanine twelve nucleotides from the 5′- end of the repeat unit. Allele 31 contains 7 repeat units not yet sequenced

Allele 14(C)	A	B	C	H	H	I	I	H	I	J	I	I	L	G													
Allele 15(AA)	A	B	C	D	E	F	C	G	H	I	J	I	I	L	G												
Allele 17(AA)	A	B	C	D	D	D	E	F	C	G	H	I	J	I	I	L	G										
Allele 17a(C)	A	B	C	D	E	C	H	H	I	I	H	I	J	I	I	L	G										
Allele 18$(C)_1$	A	B	C	D	D	E	C	H	H	I	I	H	I	J	I	I	L	G									
Allele 18$(C)_2$	A	B	C	D	D	E	C	H	H	I	I	H	I	J	I	I	L	G									
Allele 20a$(H)_1$	A	B	C	D	E	F	C	G	H	I	I	I	J	H	I	J	I	I	L	G							
Allele 20a$(H)_2$	A	B	C	D	E	F	C	G	H	I	I	I	J	H	I	J	M	I	L	G							
Allele 20a(C)	A	B*	C	D	E	F	C	G	H	I	I	I	J	H	I	J	I	I	L	G							
Allele 20m(C)	A	B	C	D	D	E	C	E	C	H	H	I	I	H	I	J	I	I	L	G							
Allele 20c(C)	A	B	C	D	D	E	C	H	H	I	I	I	I	H	I	J	I	I	L	G							
Allele 21(C)	A	B	C	D	D	E	F	C	G	H	I	I	I	J	H	I	J	I	I	L	G						
Allele 22c(H)	A	B	C	D	D	E	C	G	H	I	K	K	I	J	K	H	I	J	I	I	L	G					
Allele 22c(C)	A	B	C	D	D	D	D	E	C	H	H	H	H	I	I	H	I	J	I	I	L	G					
Allele 22c(AA)	A	B	C	D	D	D	E	N	C	G	H	I	I	I	J	H	I	J	I	I	L	G					
Allele 23c(H)	A	B	C	D	D	E	C	C	G	H	I	K	I	H	I	I	H	I	J	I	I	L	G				
Allele 24$(C)_1$	A	B	C	D	D	E	F	C	G	H	I	I	I	J	H	I	J	H	I	J	I	I	L	G			
Allele 24$(C)_2$	A	B	C	C	D	E	F	C	G	H	I	I	I	J	H	I	J	H	I	J	I	I	L	G			
Allele 25(H)	A	B	C	D	D	D	E	F	C	G	H	I	I	I	J	H	I	J	H	I	J	I	I	L	G		
Allele 25c(H)	A	B	C	P	D	E	C	C	G	H	K	K	I	K	I	H	I	I	H	I	J	I	I	L	G		
Allele 26(C)	A	B	C	D	D	E	C	E	F	C	G	O	I	I	I	J	H	I	J	H	I	J	I	I	L	G	
Allele 26a(C)	A	B	C	D	D	E	F	C	G	H	I	I	I	I	I	J	H	I	J	H	I	J	I	I	L	G	
Allele 31(C)	A	B	C	D	Q	E	F	C	G	H	I	(7RU)			I	K	I	K	I	I	H	I	R	I	I	L	G

mon to all alleles at the first repeat, there was only one example of an insertion or deletion in one 20 anodic length allele (Table 2:allele 20a(C)) where there was a deletion in the second type B repeat unit. Most D1S80 microvariation is the result of sequence variation.

Figure 1 shows the repeat unit arrays contained within selected D1S80 alleles. The most abundant repeat unit found was type I and types M–R were the rarest observed (Table 1). The H repeat unit was identical to the consensus repeat unit and the most prevalent repeat contained in two Gorilla alleles sequenced (unpublished data). The consensus sequence was determined from the frequency of occurrence of each base in a given position in the repeat unit [21]. Of the 18 core repeats 7 have an A at positions 3, 10, and 13, instead of a G and C, while repeats I–O are more varied. Positions 6, 7, 8, 12, and 16 were invariant among the repeat units studied. One example of a deletion of one nucleotide was found in a allele anodic to allele 20 which is consistent with the anodic migration.

Because of the complexity of the repeat arrays, structural differences between the alleles became more apparent when the arrays of repeat units were divided into common motifs or super-repeat domains and 6 motifs ranging from 3 to 9 repeat units were identified (Fig. 1). If the motifs differed by one or two repeat units and the repeat units were different by only one base change, they were classified as members of the same motif class.

As shown in Fig. 1, it is difficult to establish a consensus array of repeat units for alleles due to the arrangement of the motifs and the variability in the repeat number making up the length of the alleles. Several of the alleles compared with respect to repeat arrays were too diverse to evaluate an evolutionary relationship, however there are two obvious structural alleles and each has some relationship with either the 18 or the 24 allele. The motifs characterized by the 18 structural allele were, α = (A B C), Θ = (J I I L G), Γ = (D E C H H), and δ = I I H I. Alleles 14(C), 17a(C), 20m(C), 20c(C), and 22c(C) could be derived from the 18 allele by one or two mutational events (Fig. 1). For instance allele 22c(C) differs from the 18 allele by addition or loss of ‘D’ and ‘H’ repeat units (DD↔DDDD and HH↔HHHH).

The 24 structural alleles were characterized by the following motifs: α = (A B C), Θ = (J I I L G), β = (D E F C G H I) and one or several ε = (J H I) motifs. In fact one 24 allele differed from the other by only one mutational event which occurred in the fourth repeat unit from the 5′ end of the array in which a G > A transition produces a C > D repeat unit conversion. This 24 repeat allele did not migrate as an off ladder variant; it was undetected under standard conditions. Alleles 15(AA), 17(AA), 20a(H_1), 20a(H_2), 20a(C), 21(C), 22c(AA), 25(H), 26(C), and 26a(C) could be derived from the 24 allele by one or two additions or deletions of repeats or motifs (Fig. 1). For instance the 24 repeat unit allele must lose one (ε) motif and reduplicate one type D repeat unit for it to be converted to the 22c(AA) repeat unit allele which contains a (α), (β*), and one (ε) motif. The (β*) motif contains a type N repeat

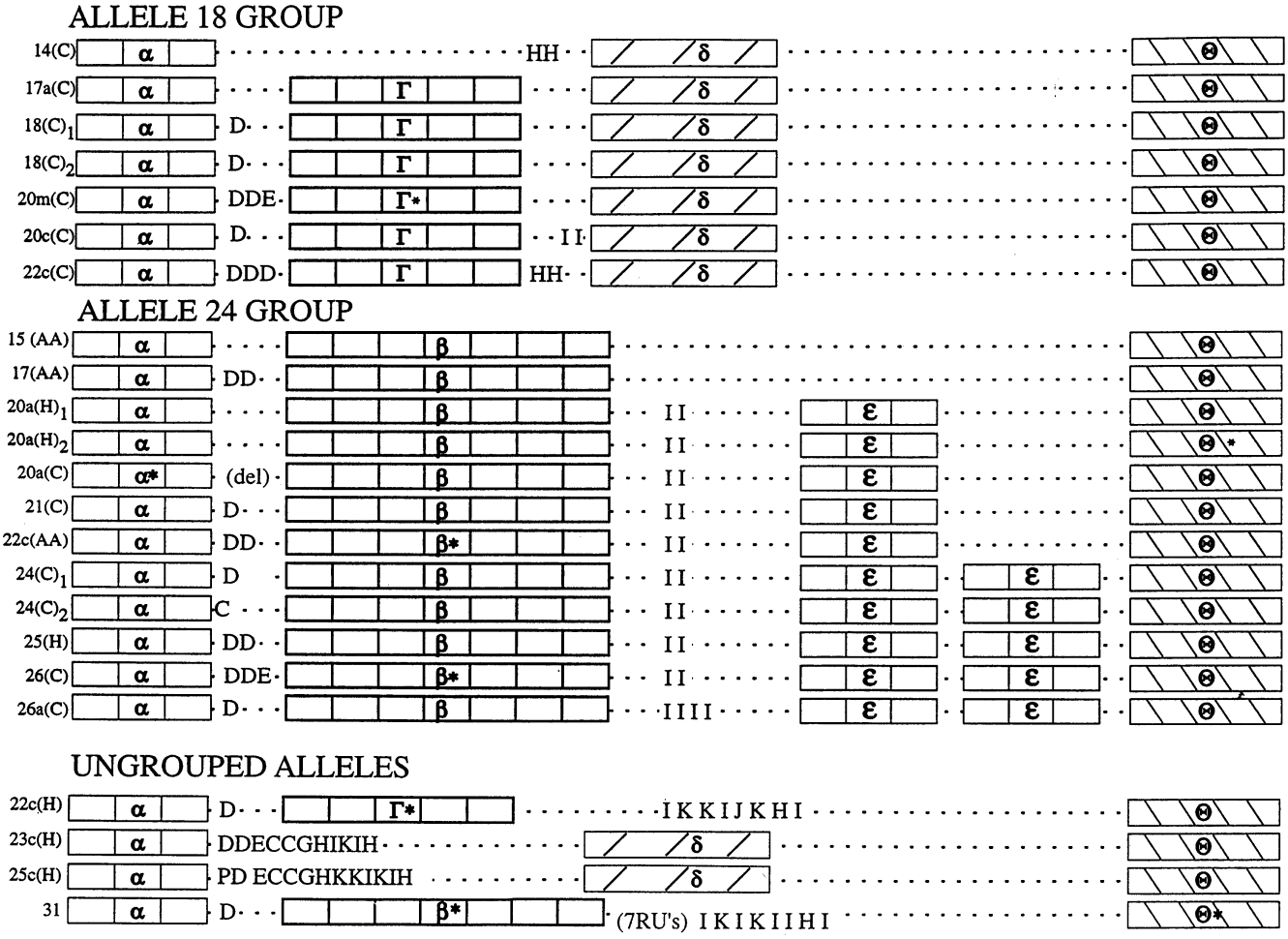

Fig. 1 Repeat unit sequence organization represented by motifs in the human D1S80 locus. The repeat length is given by the number of repeat units. The letter to the right of the number of repeat units is the position the allele falls in reference anodally (a) or cathodally (c) to the nominal allele. The letter in parentheses identifies the broad population classification of the originator of the sample; C = Caucasian, H = Hispanic, and AA = Afro American. Each repeat unit not part of a motif is represented by a Arabic letters (Table 1). Motifs are represented by greek letters and correspond to groups of repeat units as follows: α = A B C; Θ = J I I L G; β = D E F C G H I; Γ = D E C H H, δ = I I H I; ε = J H I. If the motifs differed by one or two repeat units and the repeat units were different by only one base change, they were classified as members of the same motif class and are demarcated with a (*). *(del) = one nucleotide deletion in the 'B' repeat unit of the α motif. (7RU's) = Seven repeat units that were not sequenced

unit which is one base change from an type F repeat unit which under our criteria is considered a member of the same motif class.

The above determination of the evolutionary origin of the various alleles is crude. However, there is additional support that our interpretation of motif arrangements and the associations with either the 18 or 24 ancestral alleles is correct. Alonso et al. [22] and Duncan et al. [23] described associations with restriction sites in the D1S80 flanking region and the alleles 18 and 24. There appears to be a strong association of the 5′ flanking region *Hin*f I(+) and *Tsp*509 I(–) site and the 3′ flanking region *Bso*F I(–) site with the 18 allele, while the 24 tends to be associated with the *Hin*f I(–), *Tsp*509 I(+) and *Bso*F I(+) sites. In all alleles analyzed (except 17a(C)), the restriction site haplotype polymorphisms were consistent with the assigned allele types predicted by the motif analysis.

More sequences are required to ascertain whether or not there is racial or ethnic distribution of the kinds of repeat arrangements. The data demonstrate that variation of D1S80 alleles are primarily due to the number of repeat units contained within the amplicons. Microvariation generally is due to sequence variation within repeats between alleles; However, it cannot be ruled out that some rare alleles may be due to insertions or deletions.

Acknowledgements The authors are grateful to Dr. Scott Citino and Ms Jacky Shaw of Miami Metrozoo for providing blood samples from several Gorillas. We especially thank Dr. Richard Warren, Director, Miami Children's Hospital Research Institute for his encouragement and support. This research was partially funded by a Forensic Science Foundation, Lucas Grant.

References

1. Nakamura Y, Carlson M, Krapcho K, White R (1989) Isolation and mapping of a polymorphic DNA sequence (pMCT118) on chromosome 1p (D1S80). Nucleic Acids Res 16:9364

2. Kasai K, Nakamura Y, White R (1990) Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. J Forensic Sci 35:1196–1200
3. Kadasi L, Bohusova T (1995) A new extremely large allele at the D1S80 (MCT118) locus. J Forensic Sci 40:906–907
4. Sajantila A, Budowle B, Strom M, Johnsson V, Lukka M, Peltonen L, Ehnholm C (1992) PCR amplification of alleles at the D1S80 locus:comparison of a Finnish and a north american Caucasian population sample, and forensic casework evaluation. Am J Hum Genet 50:816–825
5. Budowle B, Baechtel SF, Comey CT (1992) Some considerations for use of AMP-FLPS for identity testing. In:Rittner C, Schneider PM (eds) Advances in forensic haemogenetics, 4. Springer, Berlin Heidelberg New York, pp 11–17
6. Budowle B, Baechtel FS, Smerick JB, Presley K, Giusti AM, Parson G, Alevy MC, Chakraborty R (1995) D1S80 population data in African Americans, Caucasians, Southeastern Hispanics, Southwestern Hispanics, and Orientals. J Forensic Sci 40: 38–44
7. Baechtel FS, Smerick JB, Presley KW, Budowle B (1993) Multigenerational amplification of a reference ladder for alleles at locus D1S80. J Forensic Sci 38:1176–1182
8. Möller A., Schürenkamp M, Brinkmann B (1995) Evaluation of an ACTBP2 ladder composed of 26 sequenced alleles. Int J Legal Med 108:75–78
9. Duncan GT, Tracey ML, Kuhn D, Garrison J, Baird LS (1995) Detection of D1S80 (pMCT118) microvariant polymorphisms by single-strand conformation polymorphism analysis. Proceedings from the sixth international symposium on human identification. Promega, Madison, WI, pp 96–103
10. Möller A, Brinkmann B (1994) Locus ACTBP2 (SE33) Sequencing data reveal considerable polymorphism. Int J Legal Med 106:262–267
11. Brinkmann B, Meyer E, Junge A (1996) Complex mutational events at the HumD21S11 locus. Hum Genet 98:60–64
12. Möller A, Meyer E, Brinkmann A (1994) Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD2S11. Int J Legal Med 106:319–323
13. Buard J, Vergnaud G (1994) Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). EMBO J 13: 3203–3210
14. Jeffreys AJ, Allen MJ, Armour JAL, Collick A, Dubrova Y, Fretwell N, Guram T, Jobling M, May CA, Neil DL, Neumann R (1995) Mutation processes at human minisatellites. Electrophoresis 16:1577–1585
15. May CA, Jeffreys AJ, Armour JAL (1996) Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). Hum Mol Genet 5:1823–1833
16. Ellsworth DL, Shriver MD, Boerwinkle E (1995) Nucleotide sequence analysis of the apolipoprotein B 3′ VNTR. Hum Mol Genet 4:937–944
17. Comey CT, Koons BW, Presley KW, Smerick JB, Sobieralski CA, Stanley DM, Baechtel FS (1994) DNA extraction strategies for amplified fragment length polymorphism analysis. J Forensic Sci 39:1264–1269
18. Waye JS, Michaud D, Bowen JH, Fourney RM (1991) Sensitive and specific quantification of human genomic deoxyribonucleic acid (DNA) in forensic science specimens: casework examples. J Forensic Sci 36:1198–1203
19. Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC (1991) Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am J Hum Genet 48:137–144
20. Sanger F, Coulson AR (1975) A rapid method for determining the sequence in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:444–448
21. Lewin B (1990) Genes IV. Oxford University Press, Oxford
22. Alonso A, Martin P, Albarran C, Sancho M (1995) A Hinf I polymorphism in the 5′ flanking region of the human VNTR locus D1S80. Int J Legal Med 107:216–218
23. Duncan GT, Balamurugan K, Budowle B, Tracey ML (1996) A Hinf I/Tsp509 I and BsoF I polymorphism in the flanking region of the human vntr locus D1S80. Gen Ana Biomol Eng 13:119–121