J Forensic Sci, Jan. 2003, Vol. 48, No. 1
Paper ID JFS2001034_481
Available online at: www.astm.org

TECHNICAL NOTE

J. R. Luis,[1] Ph.D.; S. Dios,[1] Ph.D.; J. C. Carril,[1] Ph.D.; R. Herrera,[2] Ph.D.; and B. Caeiro,[1] Ph.D.

Molecular Phenotyping of a Trinucleotide Repeat (D5S373) Experimental Conditions*

ABSTRACT: The aim of this study is to assess the utility of the STR D5S373 in human identification. PCR amplification and electrophoretic separation were optimized in order to achieve unambiguous phenotyping. We concluded that primer concentration and annealing temperature are the main factors affecting the specificity of PCR. In our population survey including three human major groups (Europe, Sub-Saharan Africa, and Asia), up to six alleles and six interalleles have been found ranging in size from 86 to 101 bp. The phenotypes were determined using horizontal polyacrylamide gel electrophoresis, a technique which has turned out to be suitable for separating fragments as close as 1 bp. In each population, the genotype frequencies conformed to the expectations of genetic equilibrium. Sequence studies were carried out to make the allele nomenclature fit to ISFH recommendations. Results from our population analysis of D5S373 show clear differences in allelic frequency patterns among the three major human groups examined. Human identification parameters estimated from our study are similar to those obtained for other STRs currently used in DNA testing.

KEYWORDS: forensic science, DNA typing, polymerase chain reaction, PCR, short tandem repeats, STR, D5S373, DNA marker, genetic polymorphism

The abundance of polymorphic short tandem repeats (STRs) scattered uniformly throughout the human genome has attracted the interest of forensic geneticists to analyze the utility of these DNA markers as potential tools for human genetic identification. The number of trimeric and tetrameric loci have been estimated to be approximately 200,000 (i.e., 1 locus every 15Kb) and almost half of them are polymorphic (1). Similarly, the small size of STRs facilitate their PCR amplification, even with degraded DNA which has greatly contributed to the widespread use of these markers in forensic laboratories. Obviously, the degree of genetic polymorphism is a key factor, although this is not enough, as aspects such as ease of amplification together with molecular stability and technical reproducibility, are also major issues before approving these markers for routine analysis. This study, therefore, deals with the analysis of a new STR (D5S373), which presents a trimeric repeat sequence (TAA). It was detected in the cosmid E5.12 by hybridization with a synthetic $(TAA)_8$ oligonucleotide and is located on chromosome 5 (5q32) (2). Different experimental conditions regarding PCR amplification and molecular separation were tested, the aim of which is the development of a standard and reliable technique for the molecular diagnosis of D5S373 phenotypes. Once the morphogenetic model of this locus is established for paternity analysis, a basic goal of this study is also to investigate the degree of polymorphism for this STR in different human groups in order to assess its statistical informativeness.

[1] Department of Anthropology, Faculty of Biology, University of Santiago de Compostela, Galicia, Spain.
[2] Department of Biological Sciences, F.I.U., FL.
* This study was partially supported by grants from the Xunta de Galicia (XUGA 20011B98) and the Ministerio de Sanidad y Consumo (FIS97/1086).
Received 29 Jan. 2001; and in revised form 16 Oct. 2001, 11 June and 20 Aug. 2002; published 11 Dec. 2002.

Material and Methods

Samples were obtained from whole blood (5 mL) and treated with EDTA-Na_2 (10% w/v)(50 μL/mL of blood) as anticoagulant. DNA was obtained using the chelating resins (3) and phenol-chloroform (4) methods, and was stored at −20°C pending amplification. Primer sequences were used as described by Dixon and Dixon (2):

5′ GGT AAC AAG AGA GAA ACT CC 3′
5′ CAA TTT CTT AGT GCA CAC ATC 3′

The amplification reaction was performed in a Linus Dualcycler according to the following conditions. Temperature cycling conditions were 92°C/30 sec, 60°C/30 sec, 72°C/30 sec for 35 cycles and an additional extension step at 72°C for 10 min. The PCR was accomplished in a final volume of 12.5 μL containing 1–5 ng DNA, 0.75 μM of each primer, 200 μM dNTPs, 0.5 units of Taq DNA polymerase, 1.5 mM Mg Cl_2, 50mM KCl and Tris-HCl 10 mM pH 8.3.

PCR amplified products were separated by horizontal polyacrylamide gel electrophoresis with a discontinuous buffer system. Gel composition was 10% T, 5% C (5), using piperazine diacrylamide as a crosslinker, where T corresponds to the concentration as a percentage of both monomers (w/v), whereas C indicates the concentration as a percentage of crosslinker in the total gel composition. Electrophoresis was carried out at constant 12.5 V/cm for 2 h for 11 cm gels, and 25 V/cm for 4 h for 19 cm gels. The bands were visualized by a specific DNA Silver Staining (6).

Sequencing studies were carried out in order to assess the molecular characterization of allele variants. The DNA for sequencing was directly obtained from silver stained polyacrylamide gels and purified using the QIAquick PCR purification Kit (QIAGEN), a purification system, which combines spin-column technology and the selective binding properties of silica-gel membranes. The dye terminator cycle sequencing method was applied using the DNA Sequencing Kit (Perkin Elmer Applied Biosystems) and DNA purification was carried out by means of ethanol precipitation (7). Electrophoretic running took place in the ABI Prism™ 377 sequencer.

Statistical samples were obtained from unrelated, healthy, autochthonous individuals from the populations of Galicia (NW Spain), Benin (W Africa) and China, in an initial survey representing the main human groups from Western Europe, sub-Saharan Africa, and Eastern Asia.

Formal genetic analysis from family groups including 48 meiosis were carried out in order to infer the inheritance model. The linkage analysis between D5S373 and D5S818 loci, was assessed using the method of Long et al. (8). D5S818 phenotyping was carried out using the AmpFℓSTR® Identifiler™ PCR Amplification Kit (Applied Biosystems, Foster City, CA).

Hardy-Weinberg equilibrium was tested using the exact test based on Markov's Chain (9) which was carried out in 50 batches, each with a size of 1000, and 1000 steps as the dememorization period, using the GENEPOP program (10). In order to assess the informative yield of D5S373 statistical parameters of forensic interest, such as Expected Heterozygosity (H_e) (11), Polymorphic Information Content (PIC) (12), Mean Exclusion Chance (MEC) (13) and Discrimination Power (DP) (14), have also been estimated.

Results and Discussion

D5S373 has proven to be a locus amenable for amplification as the yield is sufficient enough for DNA detection, although too many non-specific DNA fragments are produced. Especially problematic is the presence of a constant band in the reading zone, which interferes with the correct diagnosis of phenotypes. Different experimental conditions were tested to overcome this problem. The amount of template and concentration of $MgCl_2$ did not turn out to be particularly significant. In our experience, primer concentration and annealing temperature proved to be critical factors in order to achieve a high specificity in the amplification reaction.

Primer concentration was tested in an interval range of between 2 μM and 0.25 μM. We noted that as primer concentrations decrease, the non-specific bands are increasingly faint, although this results in lower PCR yields. In order to compensate for both aspects, we fixed the optimum primer concentration at 0.75 μM, (Fig. 1*a*). However, annealing temperature is found to be the key factor. It was tested in a range of between 50–65°C. Naturally, the higher the temperature, the higher the specificity, although we simultaneously observed a decrease in the intensity of the amplified products. A compromise temperature of 60°C satisfied both aspects, since at that point, the presence of non-specific products is scarce and the amplification yield is satisfactory (Fig. 1*b*). Even so, we recommend increasing the number of cycles to 35 to achieve a more intense banding pattern.

Another relevant technical aspect is the delineation of the conditions for electrophoretic separation. The small size of the alleles of D5S373 (ranging between 86–101 bp) facilitates a better molecular separation between them as compared with other larger sized STRs. Based on our experience, for the separation of

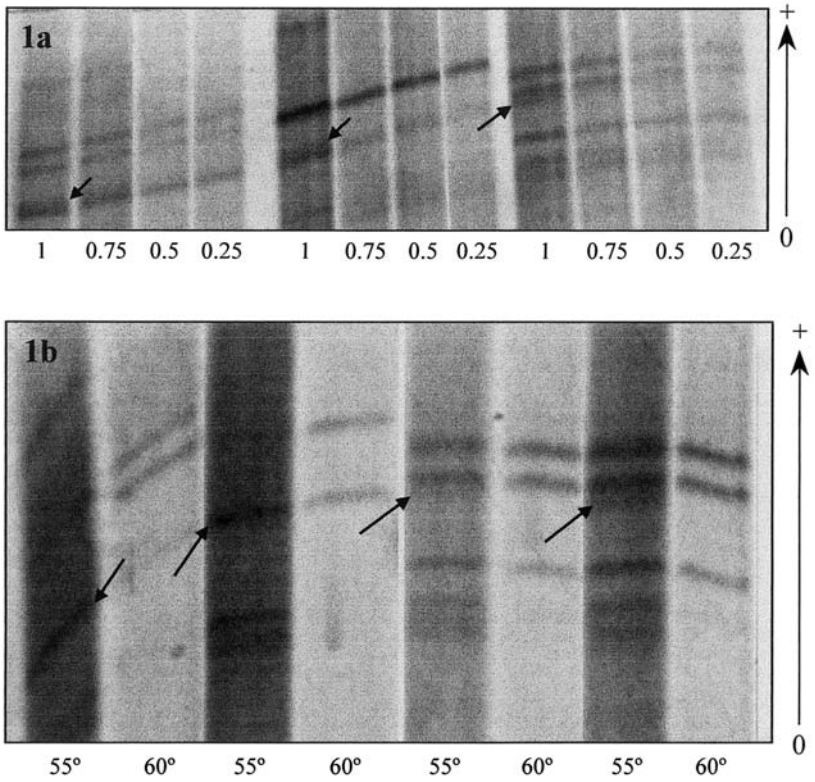

FIG. 1—*Influence of primer concentration (μM)(1a) and annealing temperature (°C) (1b), in the presence of additional and stutter bands. Arrows indicate the constant band referred to in the text.*

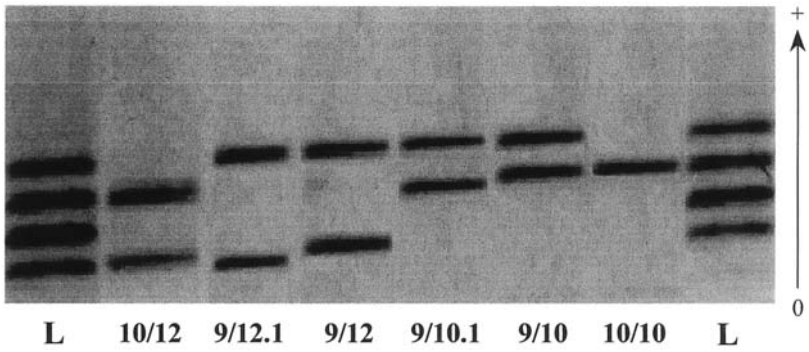

FIG. 2—*Phenotype patterns of D5S373 after electrophoresis in polyacrylamide gels, according to the optimized conditions described in the text.*

TABLE 1—*Nucleotide sequences obtained for the alleles 10.1, 10.2 and 12.1. Bold and underlined letters are referred to the nucleotide differences in relation with the consense sequence for the remaining alleles.*

10.1	GGTAACAAGAGAGAAACTCCTTCTCAT(AAT)$_{10}$ATATAT**T**ATATTCATGATGTGTGCACTAAGAAATTG
10.2	GGTAACAAGAGAGAAACTCCTTCTCAT(AAT)$_{10}$ATATAT**AT**ATATTCATGATGTGTGCACTAAGAAATTG
12.1	GGTAACAAGAGAGAAACTCCTTCTCAT(AAT)$_{12}$ATATATATATA**T**CATGATGTGTGCACTAAGAAATTG
Consense	GGTAACAAGAGAGAAACTCCTTCTCAT(AAT)$_{n}$ATATATATATTCATGATGTGTGCACTAAGAAATTG

D5S373 fragments, we recommend an acrylamide concentration of 10% and 5% of crosslinker. Molecular sieving provides a good separation of single bands for these concentrations, in horizontal polyacrilamide gels of 11 cm. This separation is further improved when electrophoresis is carried out over a longer distance (19 cm) in conjunction with a relatively higher rate of V/cm. The combination of both factors results in resolution levels high enough to differentiate interallelic variants that differ only by one pair of bases (Fig. 2).

Allelic variants were named according to the number of repeats present, taking into consideration the naming regulations recommended by the DNA Commission of the ISFH (15). Accordingly, DNA sequencing analysis was carried out for each of the fragments which have an exact number of repeats. The general structure corresponds to a design where the region of DNA found between the primers is mainly devoted to TAA trinucleotide repeats which define the length of alleles. The number of TAA repeats corresponding to the D5S373 fragments varies between 8 and 13 and results in molecular sizes ranging from 86 to 101 bp. Also noteworthy is the existence in the 5′-3′ downstream region of 5 TA dinucleotide motifs. Sequencing studies for some samples suggest that variations in this region may be responsible for the appearance of some interalleles detected in this study (Table 1), although confirmation of this issue requires further analysis. The sizes of the six interalleles observed do not correspond to a integer number of trimeric repeats, and so we have used the nomenclature recommended by the DNA Commission of the ISFH (15) to tentatively name these interalleles 9.2, 10.1, 10.2, 11.1, 11.2 and 12.1, respectively. The ladder has been constructed using alleles common in the populations under study. Using the electrophoretic conditions previously delineated for the analysis of D5S373, the separation of bands is wide enough to permit accurate identification of the molecular phenotypes including those cases involving interalleles (Fig. 2).

Table 2 displays the distribution of allelic frequencies for the populations belonging to the three main human groups. Analysis of this table reveals the utility of this locus in population characterization studies. One fact clearly illustrates this point. The most common allele is different in each of the three populations (alleles D5S373-10, 11 and 12 for Galicia, Benin and China, respectively). Allele 9 is also notable, since its frequency differs substantially among the three populations. In addition, we wish to mention allele 12.1, whose actual distribution may suggest a phylogenetic origin subsequent to the splitting of the human groups.

TABLE 2—*Allele frequencies of STR D5S373 and results of the exact test for genetic equilibrium for the populations of Galicia, Benin, and China (N = no. alleles examined).*

Alleles (nº of Repeats)	Galicia ($N = 200$)	Benin ($N = 110$)	China ($N = 86$)
8	…	0.027	0.012
9	0.145	0.291	0.023
9.2	0.005	…	…
10	0.355	0.155	0.128
10.1	…	0.009	…
10.2	…	0.027	…
11	0.275	0.400	0.326
11.1	0.005	…	…
11.2	…	0.009	…
12	0.195	0.018	0.430
12.1	…	0.055	…
13	0.020	0.009	0.081
Exact test	$p = 0.746$	$p = 0.421$	$p = 0.145$

The results of the exact test for genetic equilibrium do not show statistically significant differences between the expected and observed genotype frequencies according to the Hardy-Weinberg law (Table 2).

Analysis was conducted in 19 family groups involving a total of 48 meioses. The observed offspring phenotypic ratios from each of the 12 mating classes of this study, are concordant with a codominant single Mendelian model of transmission. Differences between observed proportions and those expected according to this model are not significant ($\chi^2 = 29.53$, $0.7 < p < 0.8$).

D5S373 is located on 5q chromosome (position: 150002556–150002650 bp; GDB Acc. no.: 19712) (2). On the other hand, another STR commonly used in paternity testing, D5S818 as well is

TABLE 3—*Values for different statistical parameters in the populations analyzed.*

	Galicia	Benin	China
H_e	0.743	0.733	0.693
PIC	0.694	0.685	0.632
MEC	0.526	0.491	0.538
DP	0.888	0.883	0.847

H_e = Expected heterozygosity.
PIC = Polymorphic information content.
MEC = Mean exclusion chance.
DP = Discrimination power.

placed on 5q chromosome (position: 126300844–126300992 bp; GDB Acc. no.: G08446) (16). In order to assess the chromosomal relationships between both STRs, linkage analyses from the genotypes of both loci have been carried out (8). The results show no statistical evidence of linkage ($p = 0.1402$), at a level of $\alpha = 0.05$.

Table 3 lists the estimated values of several forensic parameters for D5S373. It may be seen that although there are substantial differences in allelic frequency distributions among three populations studied, the numerical estimates of biostatistical parameters are similar. Both, the H_e and the MEC values for D5S373 express the biostatistical power of this marker and indicate its potential utility in paternity studies.

In summary, D5S373 displays features, which make it interesting for human genetic studies. After some modifications, PCR amplification is efficient both in specificity and productivity. The small size of amplified fragments facilitates clean electrophoretic separation of the alleles and, thus, the standardization of an accessible typing technique. Furthermore, this small size gives it high molecular stability. In conditions favorable for degradation, the number of intact copies for DNA fragments tends to be higher insofar as the locus being amplified has a lower number of base pairs. One of the major advantages of this marker is the small size of the alleles involved (86 to 101 bp) in comparison with other STR loci because a frequent problem in forensic samples is the highly fragmented state of the collected DNA. The allelic patterns among the three main ethnic groups appear to be highly differentiated, and the values of forensic statistical parameters display relatively high magnitudes, similar to those observed in other classic STRs. Furthermore, the genotype frequencies are concordant with the predictions of Hardy-Weinberg equilibrium and the inheritance pattern conforms with the Mendelian laws, so D5S373 may be used, together with other STRs, to estimate multiple loci profiles for parentage testing.

Acknowledgment

Authors are indebted to Dr. Fátima Pinheiro (Instituto de Medicina Legal do Porto, Portugal) for kindly supplying family groups for genetic analyses.

References

1. 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.
2. Dixon J, Dixon MJ. Trinucleotide repeat polymorphism at the D5S373 locus. Hum Mol Genet 1993;2:829.
3. Singer-Sam J, Tanguay RL, Riggs AD. Use of Chelex to improve the PCR signal from a small number of cells. Amplification 1989;3:11.
4. Maniatis T, Fritsch EF, Sambrook J. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982;438–54.
5. Hjertén S. Molecular sieve chromatography on polyacrylamide gels, prepared according to a simplified method. Arch Biochem Biophys 1962; Suppl 1:147–51.
6. Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC. Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am J Hum Genet 1991;48:137–44
7. Maniatis T, Fritsch EF, Sambrook J. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989;E12–E14.
8. Long JC, Williams RC, Urbanek M. An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 1995;56(3):799–810.
9. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics 1992;48:361–72.
10. Raymond M, Rousset F. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J Heredity 1995;86:248–9.
11. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89:583–90.
12. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 1980;32:314–31.
13. Krüger J, Fuhrmann W, Lichte KH, Steffens C. Zur Verwendug des Polymorphismus der sauren Erythrocytenphosphatase bei der Vaterschaftsbegutachtung. Dtsch Z Gerichtl Med 1968;64:127–46.
14. Fisher R. Standard calculations for evaluating a blood group system. Heredity 1951;5:95–102.
15. DNA recommendations—further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. Forensic Sci Int 1997;87:179–84.
16. Murray J, Sheffield V, Weber JL, Duyk G, Buetow KH. Cooperative Human Linkage Center. Unpublished 1995.

Additional information and reprint requests:
Javier R. Luis, Ph.D.
Departamento de Antropoloxía
Facultade de Bioloxía
15782 Santiago de Compostela
Galicia, Spain