Int J Legal Med (2002) 116:322–326
DOI 10.1007/s00414-002-0310-2

ORIGINAL ARTICLE

Leonor Gusmão · Cíntia Alves · Solange Costa
António Amorim · Maria Brion
Annabel González-Neira · Paula Sánchez-Diz
Angel Carracedo

Point mutations in the flanking regions of the Y-chromosome specific STRs DYS391, DYS437 and DYS438

Received: 14 December 2001 / Accepted: 4 April 2002 / Published online: 30 October 2002

Abstract Sequence analysis of the DNA fragments amplified with the DYS391, DYS437 and DYS438 primers allowed the detection of biallelic polymorphisms in the flanking region of these STR loci. In this work, we describe a methodology where both the STR alleles and the SNPs at these loci are typed. Sequencing of chimpanzee *(Pan troglodytes)* homologous loci was performed and the ancestral state of the SNPs was determined. The allele distribution of these biallelic markers was analysed within different haplogroups. For DYS391, allele 1 was found in all samples from haplogroups E3a and E* (xE3a). DYS437 allele 1 was present in all haplogroup E3a samples and absent in the haplogroup E* (xE3a). The presence of allele 1 of DYS438 was restricted to haplogroup J. The SNP typing can be helpful in distinguishing STR haplotype identity by descent from identity by state, thus proving to be very informative in forensic investigations.

Keywords Y chromosome · Biallelic marker · DYS391 · DYS437 · DYS438

Introduction

Y-chromosome biallelic markers are not currently used in forensic science, where STRs are preferred and remain the most informative approach for identification purposes. STR-defined haplotypes show a much lower degree of population stratification when compared with SNP haplogroups, which is an obvious advantage for the construction of large scale databases.

However, for STRs a significant proportion of identical haplotypes are not identical by descent due to their mutation rate. The information provided by both types of markers can therefore be very useful, at least in some forensic applications such as the distinction of identity by descent from identity by state, and the assignment of geographic origin of paternal lineage to a sample.

In this work we report an extensive study on three point mutations found in the flanking regions of the Y-STRs DYS391, DYS437 and DYS438 (Gusmão et al. 2001) by:

1. Optimisation of a methodology for detection of both the STR alleles and the biallelic marker
2. Investigation of the ancestral state of these mutations and
3. Studying samples from different haplogroups.

Materials and methods

Samples

Human blood samples, obtained under informed consent, were collected from healthy unrelated males from the following regions: 65 from Iberian Peninsula (SW Europe), 55 from Mozambique (SE Africa), 15 from Maghreb (NW Africa) and 13 from Macao (SE Asia). They were selected from previous population studies in which SNP typing was performed, in order to represent the various haplogroups across populations.

Genomic DNA was extracted as described by Valverde et al. (1993) and quantification was performed using fluorescent detection with DyNA Quant 200 (Amersham Pharmacia Biotech).

DNA samples from 10 male chimpanzees (*Pan troglodytes*) were kindly provided by Prof. Dr. W.R. Mayr from the University Clinic for Blood Group Serology and Transfusion Medicine, Department for Blood Group Serology (University of Vienna, Austria).

PCR amplification

The PCR amplification was performed using 5 ng genomic DNA in a 25 µl reaction volume comprising 1.5 mM $MgCl_2$, 1 U Taq

L. Gusmão (✉) · C. Alves · S. Costa · A. Amorim
IPATIMUP, Instituto de Patologia e Imunologia Molecular
da Universidade do Porto,
Rua Dr Roberto Frias s/n, 4200-465 Porto, Portugal
e-mail: lgusmao@ipatimup.pt,
Tel.: +351-225570700, Fax: +351-225570799

A. Amorim
Faculdade de Ciências, Universidade do Porto, Portugal

M. Brion · A. González-Neira · P. Sánchez-Diz · A. Carracedo
Institute of Legal Medicine,
University of Santiago de Compostela,
Santiago de Compostela, Spain

DNA polymerase recombinant (Gibco BRL), 1× buffer, 200 µM of each deoxyribonucleoside triphosphate (dNTP) and 0.25 µM of each forward and reverse primer. Primer sequences were those described by Kayser et al. (1997) for DYS391, and by Ayub et al. (2000) for DYS437 and DYS438. The forward primers were labelled with Cy5 (DYS391), TET (DYS437) and 6-FAM (DYS438).

Amplification was carried out in DNA thermocyclers 480 and 2400 (AB Applied Biosystems). After a 95°C pre-incubation step for 2 min, PCR amplification was performed for a total of 30 cycles using the following conditions: 94°C denaturation for 1 min, annealing at 58°C for 30 s and extension at 72°C for 30 s, followed by a 10 min final extension at 65°C.

STR typing

The PCR products were run in ABI 377 and ABI 310 genetic analysers (AB Applied Biosystems) and the ALF express II instrument (Amersham Pharmacia Biotech).

Fragment sizes were determined automatically using Genescan 2.1 analysis software (AB Applied Biosystems) and AlleleLinks version 1.00 software (Amersham Pharmacia Biotech). Samples were typed by comparison with an allelic ladder obtained using a mixture of previously sequenced samples of the most common alleles.

Alleles have been designated by the number of repeats, DYS391 according to Kayser et al. (1997) and DYS437 and DYS438 according to Gusmão et al. (2001).

STR-SNP typing

DYS391 analysis. Aliquots of 0.5 µl (5 U/µl) of the restriction enzyme Tsp45I (New England BioLabs) were mixed with 5 µl of the PCR product and 1× digestion buffer in a final volume of 10 µl. The digestion mix was incubated at 65°C for 1 h. The digested PCR products were separated by electrophoresis in an ALF express II sequencer and fragment sizes determined automatically using the AlleleLinks 1.00 software.

DYS437 analysis. Aliquots of 0.5 µl (10 U/µl) of the restriction enzyme MspI (or HpaII; MBI Fermentas) were added to 4 µl of the PCR product and 1× digestion buffer in a final volume of 5 µl. The digestion mix was incubated at 50°C for 1 h. The digested PCR products were separated by horizontal electrophoresis according to Luis and Caeiro (1995) and visualised by silver staining (Budowle et al. 1991). Genotyping was carried out by side-to-side comparison with previously sequenced samples.

DYS438 analysis. Aliquots of 0.5 µl (5 U/µl) of the restriction enzyme BseMI (MBI Fermentas) were mixed with 5 µl of the PCR product and 1× digestion buffer in a final volume of 10 µl. The digestion mix was incubated at 55°C for 1 h. The digested PCR products were separated by electrophoresis in an ABI 310 sequencer and fragment sizes determined automatically using the GeneScan 2.1 Analysis software.

DYS438 SSCP analysis

The PCR products were mixed (1:1) with formamide buffer (10 ml deonised formamide, 10 mg xylene cyanol, 10 mg bromophenol blue and 200 µl 50 mM EDTA). The samples were denatured at 95°C for 10 min, placed on ice and loaded on a 0.8× MDE gel (Mutation Detection Enhancement, BMA Bio Whittaker Molecular Applications), with 0.8× TBE and 10% glycerol.

After separation by vertical electrophoresis on a Protean II xi-cell (BioRad) apparatus, DNA fragments were visualised by silver staining (Budowle et al. 1991). Running conditions were 150 V for 10 h at 10°C.

Sequencing analysis

PCR amplified fragments were purified with Microspin S-300 HR columns (Amersham Pharmacia Biotech). A dideoxy cycling sequencing reaction was carried out using the Big Dye Terminator Cycle Sequencing Ready Reaction kit (AB Applied Biosystems). The products were purified using a $MgCl_2$/ethanol-based protocol and run in a 6% denaturing gel on an ABI 377 sequencer (AB Applied Biosystems). The results were analysed using the DNA Sequence Analysis Software 377-18.

SNP haplogroups

Haplogroups defined by 11 SNPs were typed according to Rosser et al. (2000). Samples used were included in previous reports (Rosser et al. 2000; Brion et al. 2002) and unpublished results kindly provided by Luísa Pereira (IPATIMUP), Maria de Jesus Trovoada (Departamento de Antropologia, Universidade de Coimbra; IPATIMUP) and Alexandra Lopes (IPATIMUP).

Results and discussion

Three biallelic markers were identified in the flanking regions of the DYS391, DYS437 and DYS438 STR markers. Sequence analysis of the DYS391 PCR amplified fragments demonstrated the presence of a C to G substitution at base 87 downstream of the TCTA repeat motif (Fig. 1). A base substitution (A/C) was found in DYS438, at position 7 downstream from the tandem repeat (Gusmão et al. 2001). A base substitution (C/T) at position 3 upstream from the repeat motif was also found in DYS437 (Gusmão et al. 2001).

PCR-RFLP analysis

The method in which the PCR products are enzymatically digested allows detection of both the STR and SNP alleles at the three loci under analysis.

For DYS391, the C→G substitution results in the gain of a Tsp45I restriction site that can be easily assayed by digestion with this enzyme (Fig. 1).

In the amplified fragments there are two possible recognition sites for the enzyme. A Tsp45I restriction site is present in all the samples and results in the generation of two fragments, one 99 bp in size and another varying in size between 276 and 296 bp depending on the number of repeats in the STR locus (8–13 repeats). This first restriction site has the advantage of acting as a digestion positive control. The second Tsp45I restriction site is only present in alleles where the C to G substitution is present and the digestion of the PCR products generates a pattern where the fragment is 21 bp shorter (Table 1).

The STR-SNP alleles can be typed using automatic genetic analysers (Fig. 2) or by manual electrophoresis and silver staining detection. However, if fluorescence detection is used, the dye label should be carried by the forward primer, since the Tsp45I restriction sites are located downstream of the repeat. If the reverse primer is labelled, an invariable 99 bp fragment will be detected in all the samples.

Fig. 1 Sequences of the DYS391, DYS437 and DYS438 loci (the recognition sites of the restriction enzymes Tsp45I, MspI and BseMI are in bold print)

For DYS437, the C→T substitution results in the loss of a MspI restriction site that can be assayed by digestion with this enzyme. If allele T is present, digestion of the PCR products generates a pattern in which the size fragment is 27 bp shorter (Table 1, Fig. 3). In the presence of the T→C mutation, there are no restriction sites for this enzyme and no digestion occurs.

After digestion, this marker was typed by silver staining (Fig. 3). The position of the point mutation upstream of the repeat, allows fluorescent detection when using a labelled reverse primer.

For DYS438, the A→C substitution results in the gain of a BseMI (BsrDI) restriction site. In the sequence of the DYS438 amplified fragments there are two possible recognition sites for the enzyme. One restriction site is present in all the samples thus generating two fragments upon BseMI digestion, one with 93 bp and another varying in size, depending on the number of repeats (Table 1). The second restriction site is only present in alleles where the substitution has occurred and the digestion of the PCR products generates a fragment 36 bp shorter (Table 1, Fig. 4).

Heteroduplex and SSCP analysis strategies were assayed in order to type A and C alleles. Although it is possible to find different SSCP patterns for samples carrying each of the alleles (Fig. 5), these techniques are not suitable for population screening since a large number of SSCP patterns are expected due to differences in the size and sequence structure inside the repeat (Gusmão et al. 2001).

Ancestral state of the base substitution polymorphisms

In order to determine the ancestral state of the three substitutions, the human sequences were compared with their chimpanzee homologues. From the results it was possible to conclude that the DYS391 G, DYS437 T and DYS438 C alleles occurred by mutation in the human lineages since the chimpanzee sequences always have C, C and A, respectively, at these positions.

Table 1 Fragment sizes after enzymatic digestion and allele nomenclature

Alleles	Fragment size after digestion (bp)[a]	
DYS391	Allele C(0)	Allele G(1)
8	177	156
9	181	160
10	185	164
11	189	168
12	193	172
13	197	176
DYS437	Allele C(0)	Allele T(1)
13	153	180
14	157	184
15	161	188
16	165	192
17	169	196
DYS438	Allele A (0)	Allele C (1)
9	123	87
10	128	92
11	133	97
12	138	102
13	142	107

[a]The fragment size is that expected from the sequence.

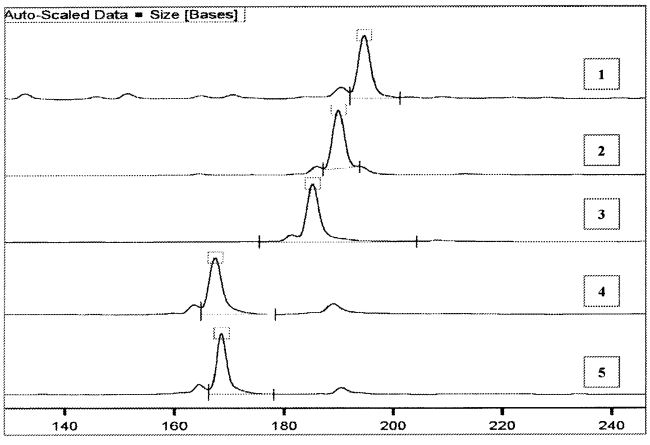

Fig. 2 DYS391 PCR/RFLP results (ALF express II, Amersham Pharmacia Biotech). STR/SNP alleles (fragment size): *lane* 1 11/0 (194.5 bp), *lane* 2, 10/0 (189.8 bp), *lane* 3, 9/0 (185.1 bp), *lanes* 4 and 5, 10/1 (167.4 and 167.8 bp)

Considering the ancestral state, the alleles were therefore named as: C(0)→G(1) for DYS391, C(0)→T(1) for DYS437 and A(0)→C(1) for DYS438.

Allele distribution

SNP allele frequencies in different population groups were studied in the context of previously defined haplogroups (according to the nomenclature of The Y Chromosome Consortium 2002).

For DYS391, a first screening was undertaken in samples already typed for other biallelic markers including

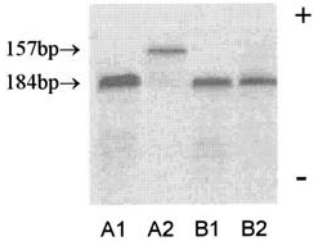

Fig. 3 DYS437 PCR/RFLP results (horizontal electrophoresis and silver staining, *A1* and *B1* STR allele 14, before digestion, *A2* sample *A1* after digestion, STR/SNP alleles 14/0, *B2* sample *B1* after digestion, STR/SNP alleles 14/1)

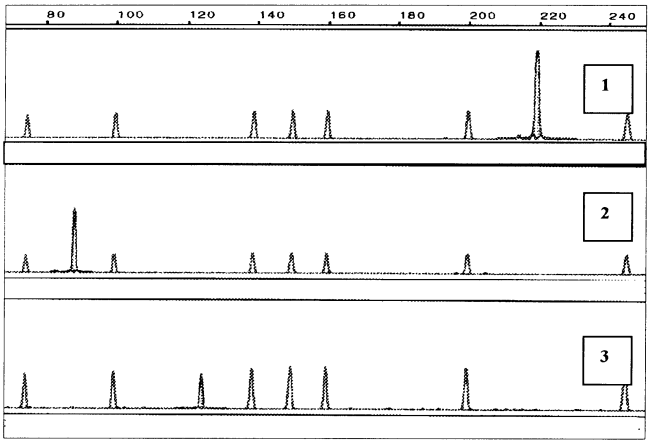

Fig. 4 DYS438 PCR/RFLP results (ABI310 genetic analyser, AB Applied Biosystems). *Lane* 1 sample before digestion, allele 10 (221 bp), *lane* 2 sample after digestion, allele 10, C(1) (92 bp), *lane* 3 sample after digestion, allele 10, A(0) (128 bp)

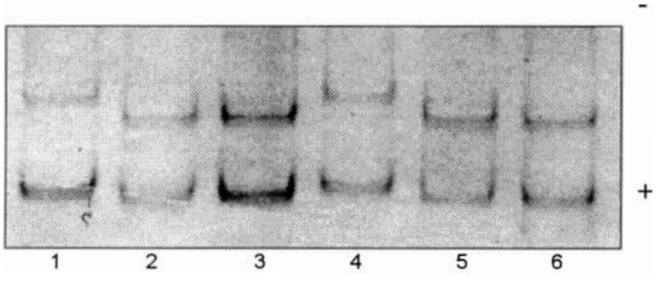

Fig. 5 DYS438 SSCP results. *Lanes* 1 and 4 STR/SNP alleles 10/0, *lanes* 2, 3, 5 and 6 STR/SNP alleles 10/1

chromosomes from haplogroups P*(xR1b8, R1a), BR*(xDE, JK), E3a, J, E*(xE3a), R1b8 and K*(xN, P) (Table 2). The ancestral state, allele C, was found in the majority of haplogroups, with the exception of HG E3a and HG E*(xE3a), where every chromosome showed the derived allele G. According to Mark Jobling (personal communication) the transversion has occurred in haplogroup E*(xE3a).

The DYS437 mutation was first detected in haplogroup E3a samples from Mozambique.

We screened for the presence of this mutation in samples from Mozambique and North Portugal, within haplogroups E3a and E*(xE3a), and allele 1 was found in all the HG E3a samples typed (Table 2).

All the haplogroup E3a and E*(xE3a) samples (n=89) carried the STR allele 14, except one haplogroup E*(E3a) sample that was allele 15. This association explains the

Table 2 Allele designation, population origin and haplogroups of the samples analysed

Haplogroup	Population group	Allele	Sample size
DYS391			
P*(xR1b8,R1a)	European	C(0)	2
BR*(xDE,JK)	European	C(0)	3
BR*(xDE,JK)	African	C(0)	3
D(xE)	Asian	C(0)	1
E3a	African	G(1)	8
J	European	C(0)	2
E*(xE3a)	European	G(1)	14
E*(xE3a)	African	G(1)	9
R1b8	European	C(0)	2
K*(xN,P)	European	C(0)	2
K*(xN,P)	Asian	C(0)	12
DYS437			
E3a	African	T(1)	49
E*(xE3a)	African	C(0)	8
E*(xE3a)	European	C(0)	32
DYS438			
BR*(xDE,JK)	European	A(0)	4
J	European	C(1)	4

very low gene diversity value found by Gusmão et al. (2001) for this STR locus in Mozambique, where allele 14 reaches a frequency of 0.969.

For DYS438, the C allele was first detected in samples from North Portugal. Based on the different SSCP patterns found, nine samples were selected for sequencing. According to the SSCP pattern, four carried the C allele and, after being typed for other biallelic markers, they were classified as belonging to haplogroup J

Conclusions

The screening of point mutations in the flanking regions of STRs is important both for technical and informative reasons. Indeed, to improve the amplification in degraded samples, it is usual in the forensic field to design new primers in order to reduce the size of the amplified fragments. For this reason, it is important to know the point mutations that could be present in order to avoid null alleles by primer mismatching (Walsh 1998; Budowle 2000; Alves et al. 2001).

In this work, methods based on PCR/RFLP analysis were developed in order to improve the applicability of these markers to large scale population surveys. These phenotyping systems proved to be very simple and allow a clear identification of the alleles both at STR and biallelic loci.

The typing of both the STR and SNP alleles increases the discrimination capacity at a specific locus and allows, at least in some cases, the distinction between identity by descent and identity by state (IBD/IBS).

Acknowledgements This work was partially supported by Fundação para a Ciência e a Tecnologia (through grant SFRH/BD/860/2000 and POCTI, Programa Operacional Ciência, Tecnologia e Inovação)

References

Alves C, Amorim A, Gusmão L, Pereira L (2001) VWA STR genotyping: further inconsistencies between Perkin-Elmer and Promega kits. Int J Legal Med 115:97–99

Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucleic Acids Res 2:e8

Brion M, Salas A, González-Neira A, Lareu MV, Carracedo A (2002) New insights into Iberian population origins through the construction of highly informative haplotypes using biallelic markers, STRs and the minisatellite MSY1. Am J Phys Anthropol (in press)

Budowle B (2000) STR allele concordance between different primer sets – a brief summary. Profiles DNA 3:10

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

Gusmão L, Alves C, Amorim A (2001) Molecular characterisation of four human Y-specific microsatellites (DYS434, DYS437, DYS438, DYS439) for population and forensic studies. Ann Hum Genet 65:285–291

Kayser M, Knijff P de, Dieltjes P, Krawczak M, Nagy M, Zerjal T, Pandya A, Tyler-Smith C, Roewer L (1997) Applications of microsatellite-based Y chromosome haplotyping. Electrophoresis 18:1602–1607

Luis JR, Caeiro B (1995) Application of two STRs (VWF and hTPO) to human population profiling. A survey in Galicia. Hum Biol 67:789–795

Rosser ZH, Zerjal T, Hurles ME, et al, (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–1543

The Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348

Valverde E, Cabrero C, Cao R (1993) Population genetics of three VNTR polymorphisms in two different Spanish populations. Int J Legal Med 151:251–256

Walsh S (1998) Commentary on: Kline MC, Jenkins B, Rogers S. J Forensic Sci 43:1103