Int J Legal Med (2006) 120: 147–150
DOI 10.1007/s00414-005-0057-7

ORIGINAL ARTICLE

Maria T. Zarrabeitia · Ana Alonso · Javier Martin · Miguel A. Gonzalez-Gay · Juan C. Martin-Escudero · Marian Martinez de Pancorbo · Pilar Sanz · Francisco Ruiz-Cabello · Jose A. Riancho

Study of six X-linked tetranucleotide microsatellites: population data from five Spanish regions

Received: 22 March 2005 / Accepted: 10 October 2005 / Published online: 19 November 2005

Abstract We studied six X-linked microsatellites in a large group of Spanish individuals (n=614) from five different regions located in northern, central and southern Spain. All the markers had tetranucleotide repeat units (DXS9895, DXS9898, DXS7130, DXS7132, GATA172D05 and DXS 6789). They were amplified in two triplex PCR reactions. There were no significant sex- or region-related differences in allelic frequencies, suggesting that general national databases can be adequate as a reference for X-linked markers. The analysis of those six short tandem repeats combined in 316 males revealed 300 different "temporary haplotypes", 283 of which were found only once. There was no evidence for statistically significant linkage disequilibrium among the loci studied. Therefore these markers are quite polymorphic and useful for forensic purposes.

Keywords STR · Microsatellites · X chromosome · Haplotypes · Multiplex

M. T. Zarrabeitia (✉) · A. Alonso
Unit of Legal Medicine, Faculty of Medicine,
University of Cantabria,
Av. Herrera Oria, S/N,
39011 Santander, Spain
e-mail: zarrabet@unican.es
Tel.: +34-942201984
Fax: +34-942201903

J. Martin
Instituto de Parasitología y
Biomedicina "Lopez Neyra", CSIC,
Granada, Spain

M. A. Gonzalez-Gay
Servicio de Reumatología,
Complejo Hospitalario Xeral-Calde,
Lugo, Spain

J. C. Martin-Escudero
Servicio de Medicina Interna,
Hospital del Río Ortega,
Universidad de Valladolid,
Valladolid, Spain

M. M. de Pancorbo
Facultad de Farmacia,
Universidad del Pais Vasco,
Vitoria, Spain

P. Sanz
Instituto Nacional de Toxicología,
Sevilla, Spain

F. Ruiz-Cabello
Servicio de Análisis Clínicos,
Hospital Universitario Virgen de las Nieves,
Universidad de Granada,
Granada, Spain

J. A. Riancho
Departamento de Medicina,
Universidad de Cantabria,
Santander, Spain

Introduction

Microsatellites of the X chromosome have been increasingly studied in recent years as a useful tool in forensic analysis. Since fathers transmit the same X chromosome to all their daughters, they are particularly useful in deficiency paternity cases when the child is a female, in maternity testing, and in paternity cases involving blood relatives [1]. Thus, several groups have published methods to amplify different microsatellites and have reported allelic frequencies [2–5]. In fact, our group reported the characterisation of five polymorphic loci on the X chromosome in a single reaction [6]. However, some trinucleotide repeats were included that are prone to show stutter bands that under some circumstances may complicate the interpretation of genetic profiles.

It has been shown that the allelic frequency distribution of autosomal short tandem repeats (STR) usually shows little variation within large populations (i.e., within a given country or even through Europe [7]). However, although frequencies from different continents have been published (for instance, see [2, 8–10]), it is unclear to which extent X-linked markers display differences at the regional level. This is an important issue in order to decide which database is needed to interpret DNA results. On the other hand, the degree of linkage disequillibrium among different markers

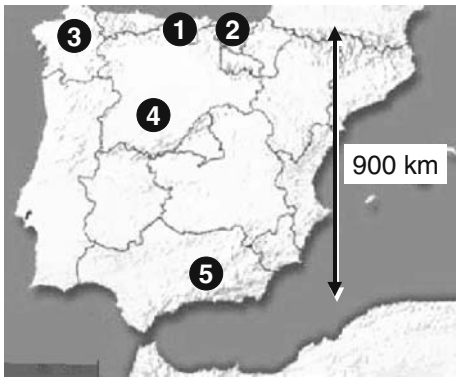

Fig. 1 Regions included in the study: *1* Cantabria, *2* Basque Country, *3* Galicia, *4* Castilla, *5* Andalucia

is unclear. This is a critical information in order to establish if the results of different loci can be treated as independent and the likelihood ratios combined.

Therefore we decided to explore the potential usefulness of different tetranucleotide microsatellites, to study the allelic frequency distribution in several populations distributed from the north through the south part of Spain, and to analyse the degree of linkage disequillibrium among the studied loci.

Materials and methods

Subjects Unrelated subjects living in five Spanish regions were studied (Fig. 1). They comprised 614 individuals, including 205 from Cantabria (75 women, 130 men), 96 from the Basque Country (52 women, 44 men), 110 from Galicia (57 women, 53 men), 115 from Castilla (57 women, 58 men), and 96 from Andalucia (57 women, 39 men).

Markers were amplified in two triplex reactions: one for markers DXS 7130, DXS 7132 and DXS 6789, and another for markers DXS9895, DXS 9898 and GATA 172D05. Previously published primer sequences were used (Table 1). Forward primers were labelled with various fluorochromes. The amounts of each primer in the reaction were as follows: 20 pmol DXS7130, 26 pmol DXS7132, 4 pmol DXS6789, 3

Table 1 Primer sequences

Marker	Primer sequences
DXS 9895 F	*FAM*-TTG GGT GGG GAC ACA GAG
DXS 9895 R	CCT GGC TCA AGG AAT TAC AA
DXS 9898 F	*HEX*-CGA GCA CAC CTA CAA AAG CT
DXS 9898 R	TCG ATT AGG TTC AGT TCC CA
GATA172D05F	*TET*-TAG TGG TGA TGG TTG CAC AG
GATA172D05R	ATA ATT GAA AGC CCG GAT TC
DXS 6789 F	*TET*-GTT GGT ACT TAA TAA ACC CTC TTT
DXS 6789 R	AAG AAG TTA TTT GAT GTC CTA TTG
DXS 7130 F	*FAM*-CTG CAA GCC ATT TGG AAT AT
DXS 7130 R	TCC TAG GAC TGG GAA AGG AC
DXS 7132 F	*HEX*-GAG CCC ATT TTA ATA AAT CC
DXS 7132 R	GCC AAA CTC TAT TAG TCA AC

F Forward, *R* reverse

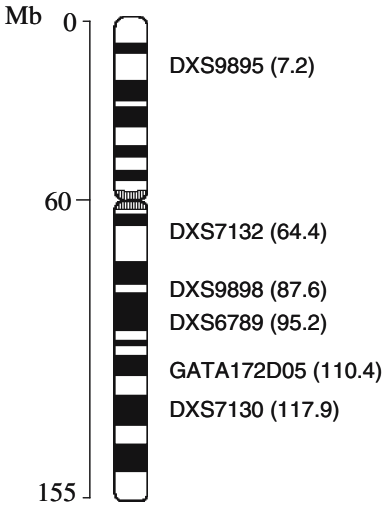

Fig. 2 Location of STRs on the X chromosome

pmol DXS9898, 6 pmol DXS9895 and 3 pmol GATA172 D05. The total reaction volume was 25 μl and contained 1.5 mM $MgCl_2$, 5 nmol each dNTP, and 2 U *Taq* polymerase. The amplification protocol consisted of an initial denaturation at 94°C for 10 min, followed by 8 cycles with denaturation at 94°C for 1 min, annealing at decreasing temperature between 62 and 59°C (1°C decrease every 2 cycles), and extension at 72°C for 1 min. Then 24 cycles at 94°C for 1 min, 58°C for 1 min, 72°C for 1 min, followed by a final extension at 72°C for 30 min.

Aliquots containing 1.2 μl of PCR product were mixed with 24 μl formamide and 1 μl TAMRA as an internal standard, heated for 5 min at 95°, quenched in an ice bath for 10 min and injected into an ABI prism 310 capillary electrophoresis system (Applied Biosystems, Foster City, CA, USA). Fragment size was determined with the local

Table 2 Allelic frequencies (triplex 1)

	DXS9895		DXS9898		GATA172D05	
	Frequency	SE	Frequency	SE	Frequency	SE
6	–	–	–	–	0.162	0.012
7	–	–	–	–	0.004	0.002
8	–	–	–	–	0.170	0.012
8.3	–	–	0.268	0.015	–	–
9	–	–	–	–	0.060	0.008
10	–	–	0.011	0.003	0.287	0.015
11	–	–	0.208	0.013	0.204	0.013
12	0.008	0.003	0.290	0.015	0.111	0.010
13	0.273	0.015	0.160	0.012	0.001	0.001
14	0.196	0.013	0.056	0.008	–	–
15	0.340	0.016	0.006	0.003	–	–
16	0.150	0.012	–	–	–	–
17	0.030	0.006	–	–	–	–
18	0.003	0.002	–	–	–	–
Between regions (*p*)	0.89		0.80		0.52	
Fst	0.004		0.003		0.002	

Table 3 Allelic frequencies (triplex 2)

	DXS7130		DXS7132		DXS6789	
	Frequency	SE	Frequency	SE	Frequency	SE
9	0.001	0.001	0.001	0.001	–	–
10	0.005	0.002	–	–	–	–
11	0.026	0.005	0.018	0.004	0.001	0.001
12	0.088	0.009	0.114	0.010		
13	0.039	0.006	0.289	0.015	0.001	0.001
13.3	0.034	0.006	–	–	–	–
14	–	–	0.326	0.015	0.010	0.003
14.3	0.244	0.014	–	–	–	–
15	–	–	0.190	0.013	0.041	0.006
15.3	0.427	0.016	–	–	–	–
16	–	–	0.050	0.007	0.013	0.004
16.3	0.118	0.011	–	–	–	–
17	–	–	0.006	0.003	0.003	0.002
17.3	0.015	0.004	–	–	–	–
18	–	–	–	–	0.004	0.002
18.3	0.001	0.001	–	–	–	–
19	–	–	0.001	0.001	0.036	0.006
19.3	0.002	0.001	–	–	–	–
20	–	–	0.002	0.001	0.382	0.016
21	–	–	0.002	0.001	0.263	0.014
22	–	–	–	–	0.156	0.012
23	–	–	0.001	0.001	0.066	0.008
24	–	–	–	–	0.024	0.005
Between regions (p)	0.035		0.67		0.15	
Fst	0.066		0.006		0.002	

Southern method implemented in Genescan software. Some common alleles were sequenced as previously reported to help in allele typing [11]. We also typed DNA sample NA9947 (available in the Poweplex kit from Promega, Madison, WI, USA) as a control. The results were in agreement with those reported by Szibor et al. [12].

DNA analysis Six STRs with tetranucleotide repeat units were studied (Fig. 2). Genomic DNA was extracted from peripheral blood by a commercial method, according to manufacturer's instructions (Qiagen, Hilden, Germany), and quantified by light absorbance. An aliquot containing 10–30 ng DNA was used to amplify the regions of interest by PCR.

Data analysis Sex and population differences in allelic frequencies were tested by a Monte Carlo extension of Fisher's exact test, using SPSS software (SPSS Inc., Chicago, IL, USA). Disequilibrium linkage and Wright's Fst values for population stratification were estimated with GDA software (P.O. Lewis and D. Zaykin, Genetic Data Analysis, v.1.1. http://lewis.eeb.uconn.edu/lewishome/software.html), which implements the formulae published by Weir [13]. Locus and haplotype diversities were calculated as $(1-\sum p_i^2)(n/n-1)$, where p_i is the allele or haplotype frequency [14]. The matching probability was estimated as 1−diversity. The polymorphism information content (PIC) was estimated as $1-\sum pi^2-\sum\sum pi^2pj^2$. The average power of discrimination (PD) and the expected probability of exclusion (PE) were estimated with the following formulae proposed by Desmarais et al. [2].

$$\mathrm{PD\ (female\ sample)} = 1-2\left(\sum p_i^2\right)^2+\sum p_i^4$$

$$\mathrm{PD\ (male\ sample)} = 1-\sum p_i^2$$

$$\mathrm{PE\ (trio)} = 1-\sum p_i^2+\sum p_i^4-\left(\sum p_i^2\right)^2$$

$$\mathrm{PE\ (motherless)} = 1-2\left(\sum p_i^2\right)+\sum p_i^3$$

Results

There were no sex-related differences in allelic frequencies. Likewise, the frequency distributions were very similar in all the populations studied. Therefore, combined data from male and female individuals from all regions studied are shown in Tables 2 and 3. Only locus DXS7130 displayed some between-region differences, marginally significant in the unadjusted analysis (p=0.035), but not significant after Bonferroni adjustment for multiple comparisons. The overall Fst value was 0.009.

Data about gene diversity and the forensic efficiency of genetic markers are shown in Table 4.

The hemizygosity of male subjects allows an analysis of temporary haplotypes to be performed. Thus in 316 male subjects, 300 different haplotype-like allelic combinations were found, 15 were observed twice, one was found in three subjects, and 283 were found only once. Therefore, haplotype-like diversity was 0.9996, and the resulting matching probability was 0.0004.

We did not find definitive evidence for linkage disequilibrium among the loci studied. When data from the hemizygotic males were analysed, all but one of the loci pairs tested for linkage disequillibrium were associated with p values well above 0.15. Only the pair DXS7130–GATA1 72D05 (separated by 7.5 Mb) was associated with a significant crude p value of 0.032. However, it was still far away from the p value

Table 4 Forensic efficiency of the markers studied

	DXS7130	DXS7132	DXS6789	DXS9895	GATA172D05	DXS9898
PIC	0.716	0.739	0.816	0.727	0.790	0.753
PD female	0.894	0.903	0.904	0.894	0.934	0.911
PD male	0.733	0.758	0.752	0.748	0.805	0.772
PE trio	0.699	0.719	0.718	0.705	0.777	0.735
PE motherless	0.560	0.583	0.582	0.566	0.653	0.601

PIC Polymorphism information content, *PD* power of discrimination, *PE* power of exclusion

of 0.002 that represents the significance threshold after Bonferroni adjustment for multiple comparisons. Furthermore, when data from females were analysed, we did not find evidence for significant disequillibrium between those two loci (p=0.41). In fact, in female samples only two loci pairs were associated with crude p values below 0.05, but still above the 0.002 threshold: DXS9895–DXS9898 (p=0.006), and DXS9895–DXS6789 (p=0.01). However, there was no evidence for disequillibrium between those loci and DXS7132, which has an intermediate location (see Fig. 2).

Discussion

The six X-linked markers reported in this paper can be amplified in two PCRs and may be quite useful in forensic cases. DXS7130 and DXS9895 were somewhat less informative than the other markers. As in the case with autosomal markers, these X-linked loci did not reveal a significant population stratification at the level studied here, which included individuals from the northern, central and southern parts of Spain. Therefore, it seems reasonable to use general national reference databases when interpreting the results.

The human X chromosome sequence has been recently elucidated, comprising about 155 Mb [15]. Markers analysed in this study are distributed throughout a 100-Mb region. A note of caution is usually raised about the lack of independence when interpreting the results of several markers located on the same chromosome. Indeed, some linkage disequilibrium among several X-linked markers was previously reported [16]. However, despite being theoretically possible, in the present study we did not find clear statistical evidence for such a disequillibrium. It should be stressed that physical linkage is not the only cause of an apparent linkage disequilibrium. Other phenomena, including population stratification, can have the same result. On the other hand, as a rule of thumb, the genetic distance of 1 cM (1 recombination per 100 meiosis) roughly corresponds to a physical distance of 1 Mb. However, there is now evidence that recombination rates vary largely throughout the genome and is not just a monotonic function of the distance. It is recognised that factors other than the physical distance between loci are involved [15, 17]. Indeed, the X chromosome appears to be rather prone to recombination, and an absence of linkage has been reported in other studies, even when quite close markers were studied [8, 18].

Nevertheless, even if the results are conservatively interpreted at the "haplotype" level, because of the theoretical possibility of linkage disequilibrium, which could potentially be demonstrated in larger studies, these markers appear to be quite discriminating. X chromosomes may recombine during female, but not male, meiosis. Thus, fathers transmit the same single X chromosome to all their daughters. Therefore, it is interesting to know the frequency and the discriminative power of those "temporary haplotypes" present in men. In the present study, we found 283 unique haplotypes among 316 males, with a matching probability of 0.04%.

In conclusion, this set of tetranuclotide X-linked STRs appears to be quite useful in forensic cases. They do not show clear evidence for disequilibrium linkage and are quite discriminating; moreover, general databases seem to be adequate for reference purposes.

Acknowledgement This study was supported by a grant from Fundación Marqués de Valdecilla-IFIMAV.

References

1. Szibor R, Krawczak M, Hering S, Edelman J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74
2. Desmarais D, Zhong Y, Chakraborty R, Perreault C, Busque L (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). J Forensic Sci 43:1046–1049
3. Hering S, Szibor R (2000) Development of the X-linked tetrameric microsatellite marker DXS9898 for forensic purposes. J Forensic Sci 45:929–931
4. Hering S, Kuhlisch E, Szibor R (2001) Development of the X-linked tetrameric microsatellite marker HumDXS6789 for forensic purposes. Forensic Sci Int 119:42–46
5. Tun Z, Honda K, Nakatome M et al (1999) Simultaneous detection of multiple STR loci on sex chromosomes for forensic testing of sex and identity. J Forensic Sci 44:772–777
6. Zarrabeitia MT, Amigo T, Sañudo C, Zarrabeitia A, González-Lamuño D, Riancho JA (2002) A new pentaplex system to study short tandem repeat markers of forensic interest on X chromosome. Forensic Sci Int 129:85–89
7. Henke L, Aaspollu A, Biondo R et al (2003) Evaluation of the STR typing kit Powerplex 16 with respect to technical performance and population genetics: a multicenter study. In: Brinkmann B, Carracedo A (eds) Progress in Forensic Genetics 9. Elsevier, Amsterdam, pp 789–794
8. Lee HY, Park MG, Jeong CK et al (2004) Genetic characteristics and population study of 4 X-chromosomal STRs in Koreans: evidence for a null allele at DXS9898. Int J Legal Med 118:355–360
9. Szibor R, Lautsch S, Plate I, Beck N (2000) Population data on the X chromosome short tandem repeat locus HumHPRTB in two regions of Germany. J Forensic Sci 45:231–233
10. Jia Y, Zhou B, Liang WB, Ly ML, Zhang L (2004) Two X-chromosome STR loci DXS6803 and XS6793 frequency data in Chinese population. J Forensic Sci 49:845–846
11. Zarrabeitia MT, Amigo T, Sañudo C, Martinez MA, Riancho JA (2002) Sequence structure and population data of two novel X-linked markers: DXS7423 and DXS8377. Int J Legal Med 116:368–371
12. Szibor R, Edelman J, Hering S et al (2003) Cell line DNA typing in forensic genetics—the necessity of reliable standards. Forensic Sci Int 138:37–43
13. Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland
14. Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
15. Ross MT, Grafham DV, Coffey AJ et al (2005) The DNA sequence of the human X chromosome. Nature 434:325–337
16. Edelman J, Szibor R (2003) The X-linked STRs DXS7130 and DXS6803. Forensic Sci Int 136:73–75
17. Taillon-Miller P, Bauer-Sardina I, Saccone NL et al (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 25:324–328
18. Hering S, Brundirs N, Kuhlisch E et al (2004) DXS10011: studies on structure, allele distribution in three populations and genetic linkage to further q-telomeric chromosome X markers. Int J Legal Med 118:313–319