Int J Legal Med (2007) 121:192–197
DOI 10.1007/s00414-006-0144-4

ORIGINAL ARTICLE

Genetic diversity of 10 X chromosome STRs in northern Portugal

Rui Pereira · Iva Gomes · António Amorim · Leonor Gusmão

Received: 17 July 2006 / Accepted: 27 November 2006 / Published online: 6 January 2007

Abstract Genetic data of 10 X chromosome STRs (DXS8378, DXS9898, DXS8377, HPRTB, GATA172D05, DXS7423, DXS6809, DXS7132, DXS101 and DXS6789) were obtained in a sample of unrelated males born in the five northern Portuguese districts. In a global sample of 347 individuals, no shared haplotypes were found for this set of markers and single locus gene diversities were high, varying between 0.678 for DXS7423 and 0.921 for DXS8377. Linkage disequilibrium analysis did not reveal consistent evidence of association between the X-STRs used. Population comparisons of northern Portuguese districts (exact test of population differentiation; pairwise genetic distances) and analysis of molecular variance supported genetic homogeneity of this region and therefore a common genetic database was considered. In comparisons with other European data, the only population samples showing statistically significant differences to northern Portugal were Germany and Latvia. The present work demonstrates that these genetic markers are highly discriminating and therefore useful for human identification purposes and anthropological research.

Keywords X chromosome · STR · Genetic identification · Population data · Northern Portugal

Electronic supplementary material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00414-006-0144-4 and is accessible for authorized users.

R. Pereira · I. Gomes · A. Amorim · L. Gusmão (✉)
Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP),
Rua Dr. Roberto Frias, s/n,
4200-465 Porto, Portugal
e-mail: lgusmao@ipatimup.pt

I. Gomes
Unidad de Genética Forense, Instituto de Medicina Legal,
Universidad de Santiago de Compostela,
15782 Santiago de Compostela, Spain

A. Amorim
Faculdade de Ciências da Universidade do Porto,
4050 Porto, Portugal

Introduction

The potential of genetic markers present on autosomes, Y chromosome and mtDNA was extensively explored in the study of genetic diversity of human populations and individual identification. Nevertheless, a major contribution of the X chromosome is still missing because only a limited number of well-characterized X-linked polymorphisms was described and population data are scarce [28]. Interest in the X chromosome has increased in recent years, mainly because it combines desirable features of other commonly used genetic markers: as with the autosomes, the X chromosome recombines and, similarly to mtDNA and Y chromosome, has a sex-biased mode of inheritance and allows direct haplotyping. Distinctive genetic features are, therefore, expected for the X chromosome in comparison with autosomal markers [22], namely, (1) a lower genetic diversity due to less frequent mutational events in females than in males [13, 15] and smaller effective population size, (2) a more pronounced genetic structure due to reduced effective population size and consequent stronger genetic drift and (3) stronger linkage disequilibrium because the X chromosome recombines only in females.

These features make the X chromosome a singular source of information in population genetics and anthropological research [12, 22] and an important tool in forensic cases [1, 21, 24, 25]. X chromosome analysis can efficiently complement the study of other genetic markers,

namely, in paternity tests involving female offspring, absence of putative fathers or blood relatives as putative fathers and in mother–son kinship testing. In these particular situations, X chromosomal STRs present higher mean exclusion chances (MECs) than autosomal STRs.

In these kinds of studies, a large amount of data concerning population profiling is needed. For the Portuguese population, although a large amount of information is already available regarding genetic variation of autosomal [10, 19], Y chromosome [2] and mtDNA markers [18], there are no available data on the X chromosome genetic diversity and population structure.

The present work aimed at the genetic characterization of 10 X-linked markers in the five administrative districts from northern Portugal (Porto, Braga, Viana do Castelo, Vila Real and Bragança). We also intended to assess the population structure in this region, to study the association between these X-STR markers and to evaluate their forensic efficiency.

Materials and methods

DNA samples

After informed consent, blood samples were collected from 347 unrelated males born in 1 of the 5 Portuguese administrative districts located north of the Douro river: Porto (n=100), Braga (n=73), Viana do Castelo (n=64), Vila Real (n=54) and Bragança (n=56). Genomic DNA was extracted from bloodstains using the Chelex method according to Lareu et al. [14].

Marker typing

The 10 X-STRs (DXS8378, DXS9898, DXS8377, HPRTB, GATA172D05, DXS7423, DXS6809, DXS7132, DXS101 and DXS6789) were typed according to the methodology of Gomes et al. [9]. A new allele found at the DXS7423 locus was sequenced using the same methodology as in Gomes et al. [9].

Statistical analysis

Allele frequencies, haplotype and gene diversities, exact test of population differentiation population pairwise genetic distances F_{ST} and R_{ST}, analysis of molecular variance (AMOVA) and pairwise exact test of linkage disequilibrium were all performed using the Arlequin ver. 3.0 software [8].

Genetic distance analysis was performed for all X-STRs assuming the stepwise mutation model (R_{ST}; [23]) with the exception of DXS9898. Because this marker presents a very frequent non-consensus allele 8.3 due to an insertion/deletion rather than a single step mutation, depending on how the allele is coded in statistical analysis software (i.e. the number of repeats considered for that allele), very different genetic distances are obtained. Therefore, for DXS9898, genetic distances were based on the number of different alleles (F_{ST}). Forensic efficiency parameters were assessed for each locus following Desmarais et al. [4].

Results

Genetic variation

In the global sample of 347 individuals from northern Portugal, when studying the 10 X-STRs simultaneously, no shared haplotypes were found and consequently the haplotype diversity is 1. Allele frequencies and gene diversity (same as the power of discrimination in males; PD_M) observed for each marker are presented in Table 1. An allele not previously described was found in the DXS7423 locus with a size corresponding to 10 repeats. Sequence analysis of this new allele showed the same repeat structure previously reported by Edelmann et al. [6]. In the HPRTB locus, an allele 13 (D48AGdel) was also found having the same sequence structure described by Gomes et al. [9].

Comparisons of the five northern Portuguese districts

Allele frequencies and gene diversities observed for the 10 X-STRs in each district are available in Table S1. Since the presence of population structure in northern Portugal would implicate the use of different databases for each subpopulation, we performed comparisons between the five districts, consisting of pairwise population comparisons at the single locus (exact test of population differentiation and genetic distance analysis) and haplotype levels.

When considering both methodologies, single locus analyses only presented significant values after Bonferroni correction for DXS9898, in comparisons between Bragança and Porto or Braga (complete results in Table S2). When taking into account haplotypes obtained by studying the 10 X-STRs, pairwise F_{ST}s were low, yet with significant p values in comparisons between Braga vs Vila Real and Bragança. Nevertheless, to bypass the aforementioned limitations caused by codification of non-consensus repeats in the assessment of genetic distances, we excluded DXS9898 from the analysis and assumed the stepwise mutation model for the remaining markers because it is more appropriate for microsatellite loci. In this case, no significant R_{ST}s were found between the five districts.

Table 1 Allele frequencies and forensic efficiency parameters for the 10 X-STRs in northern Portuguese population

	DXS8378		DXS9898		DXS8377		HPRTB		GATA172D05		DXS7423		DXS6809		DXS7132		DXS101		DXS6789	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	9	0.0144	8.3	0.2651	38	0.0029	10	0.0029	6	0.2017	10	0.0029	27	0.0086	11	0.0202	14	0.0029	15	0.0519
	10	0.2795	10	0.0115	39	0.0115	11	0.0115	7	0.0029	13	0.0605	28	0.0202	12	0.1009	15	0.0202	16	0.0231
	11	0.3055	11	0.2248	40	0.0115	12	0.1441	8	0.1671	14	0.3573	29	0.0202	13	0.2911	16	0.0029	19	0.0202
	12	0.3689	12	0.2853	41	0.0202	13[a]	0.0086	9	0.0375	15	0.4150	30	0.0490	14	0.3545	17	0.0058	20	0.4352
	13	0.0288	13	0.1527	42	0.0403	13	0.3285	10	0.3199	16	0.1412	31	0.2017	15	0.2017	18	0.0836	21	0.2305
	14	0.0029	14	0.0548	43	0.0288	14	0.3112	11	0.1988	17	0.0231	32	0.1758	16	0.0288	19	0.0461	22	0.1354
			15	0.0058	44	0.0576	15	0.1441	12	0.0720			33	0.2709	17	0.0029	20	0.0173	23	0.0836
					45	0.0432	16	0.0403					34	0.1585			21	0.0288	24	0.0144
					46	0.0749	17	0.0086					35	0.0720			22	0.0288	25	0.0029
					47	0.0980							36	0.0202			23	0.0922	26	0.0029
					48	0.1009							37	0.0029			24	0.1729		
					49	0.1239											25	0.1527		
					50	0.1210											26	0.1527		
					51	0.1153											27	0.1268		
					52	0.0317											28	0.0346		
					53	0.0519											29	0.0259		
					54	0.0231											30	0.0058		
					55	0.0144														
					56	0.0173														
					57	0.0115														
MEC_T	0.6295		0.7337		0.9123		0.7121		0.7503		0.6173		0.7976		0.6933		0.8747		0.6935	
MEC_D	0.4834		0.5995		0.8444		0.5752		0.6198		0.4721		0.6800		0.5534		0.7868		0.5545	
PD_M	0.6934		0.7735		0.9208		0.7539		0.7852		0.6779		0.8234		0.7397		0.8880		0.7304	
PD_F	0.8429		0.9101		0.9875		0.8987		0.9203		0.8364		0.9446		0.8869		0.9761		0.8914	

MEC_T: mean exclusion chance in trios involving daughters, *MEC_D*: mean exclusion chance in duos father/daughter, *PD_F*: power of discrimination in females, *PD_M*: power of discrimination in males.

[a] Indicates allele 13 (D48AGdel).

Furthermore, the AMOVA considering all studied X-linked markers also revealed that genetic variation among districts is very low (F_{ST}=0.0056; p=0.0007) and similar results were obtained when assuming the stepwise mutation model (excluding DXS9898) as the overall R_{ST} was 0.0014 (p=0.3364).

Linkage disequilibrium analysis

Linkage disequilibrium tests for all pairs of loci did not reveal consistent evidence of association between the studied markers. After Bonferroni correction for multiple analyses, we only obtained a significant p value for one pair of loci: DXS9898–DXS6789 in Viana do Castelo (p=0.0005). However, when considering our global sample from northern Portugal, the pairwise LD test did not confirm the existence of association between these markers.

Forensic efficiency parameters

Forensic efficiency statistical parameters were calculated for each locus in the northern Portuguese sample and are presented in Table 1. In our study, DXS7423 was the least polymorphic marker whereas the most diverse was DXS8377. Overall values obtained for the power of discrimination were high in females (PD_F= 0.999999999996), and males (PD_M=0.9999999) and for the combined mean exclusion chances in trios (MEC_T= 0.9999994) and duos (MEC_D=0.99996).

Comparison with other European populations

Genetic distances were calculated between northern Portuguese and other European populations with available data for the same markers studied in our sample [3, 5, 7, 16, 17, 20, 26, 29–34]. In most comparisons, low non-significant genetic distances were obtained (Table S3). Nevertheless, significant genetic distances were found for DXS8378 in comparison with Germany and Latvia and for DXS8377 in comparison with Spain, Italy, Germany or Austria.

Discussion

The present study shows that there is a high genetic diversity in the northern Portuguese population for the 10 studied X-STRs. Comparisons between the five northern Portuguese districts did not reveal a clear sub-structure of the population and, furthermore, the AMOVA results confirmed that there was no significant genetic variation between districts. Therefore, data obtained so far indicate no genetic heterogeneity in northern Portugal considering the X-STRs used in this work and so a single genetic database can be established. This is in line with previously reported results regarding genetic variation of the Y chromosome for this region [2].

Concerning linkage analysis, a higher LD could be expected between X-linked markers in comparison with autosomal markers, as the recombination rate of the X chromosome is approximately two-thirds lower, and consequently LD decays more slowly over generations. In our study, although one significant value was found for DXS9898 and DXS6789 in Viana do Castelo, pairwise LD tests for the global sample did not confirm the existence of association between these two markers. On the other hand, they are located relatively distant on the chromosome (9.3 cM) [24] and so if an association existed between any loci used, it would be expected to show up for loci pairs closely located, as for example DXS6809 and DXS6789 (see haplotype frequencies in Table S4), which were previously reported to be in strong linkage disequilibrium [27].

In general, the X-STRs used in this work proved to be highly discriminating and therefore useful for forensic purposes. Overall values of the power of discrimination were high, supporting the potential of this decaplex in identification studies. The high values obtained for the combined mean exclusion chance also confirm the utility of this system in kinship testing, namely, in paternity tests involving female offspring when complete trios or only alleged father/daughter duos are available, in paternity cases implicating a father and his son as putative fathers and in maternity investigations of male descendents.

When comparing our sample with other European samples, we experienced some difficulties as only a small subset of the studied markers in those populations were coincident with our 10 X-STRs. The comparisons performed revealed significant differences to German and Latvian populations, presenting significant, high R_{ST} values for DXS8378 even after Bonferroni correction for multiple analyses. In addition, concerning DXS8377, R_{ST} values were abnormally high between any populations from two different groups (Portugal, Latvia and Finland vs Spain, Italy, Germany and Austria), but low between populations within each group. However, these unexpected results could have other explanations than population heterogeneity.

The present study raised several questions concerning the assessment of genetic distances (mainly due to the presence of non-consensus alleles), which somehow restricted the analysis of population structure and evaluation of genetic diversity. For instance, due to the high frequency of non-consensus alleles at DXS9898, it was not possible to assess genetic distances based on the stepwise mutation model for this marker, which limits its use in this context. With the growing availability of X-STRs for

markers with similar diversities, preference should be given to those with a simple structure and absence of non-consensus repeats to better explore the informative power of the X chromosome both in forensic and population genetic studies according to the ISFG guidelines presented for Y-STRs [11]. On the other hand, the establishment of an appropriate set of markers would be of great utility as analogous haplotypic information could be used in the assessment of the sub-structure and history of populations.

In conclusion, apart from the discussed limitations of DXS9898, the X-linked markers used in this work proved to be highly informative and an important tool for human identification purposes and general anthropological research.

Acknowledgments The authors would like to acknowledge Cíntia Alves for her collaboration. This work was partially supported by the Fundação para a Ciência e a Tecnologia (through grant SFRH/BD/21647/2005 and POCI, Programa Operacional Ciência e Inovação 2010).

References

1. Asamura H, Sakai H, Kobayashi K, Ota M, Fukushima H (2006) MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis. Int J Legal Med 120:174–181
2. Beleza S, Gusmão L, Lopes A et al (2006) Micro-phylogeographic and demographic history of Portuguese male lineages. Ann Hum Genet 70:181–194
3. Bini C, Ceccardi S, Ferri G et al (2005) Development of a heptaplex PCR system to analyse X-chromosome STR loci from five Italian population samples. A collaborative study. Forensic Sci Int 153:231–236
4. 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
5. Edelmann J, Hering S, Michael M et al (2001) 16 X-chromosome STR loci frequency data from a German population. Forensic Sci Int 124:215–218
6. Edelmann J, Deichsel D, Hering S, Plate I, Szibor R (2002) Sequence variation and allele nomenclature for the X-linked STRs DXS9895, DXS8378, DXS7132, DXS6800, DXS7133, GATA172D05, DXS7423 and DXS8377. Forensic Sci Int 129:99–103
7. Edelmann J, Deichsel D, Plate I, Kayser M, Szibor R (2003) Validation of the X-chromosomal STR DXS6809. Int J Legal Med 117:241–244
8. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47–50
9. Gomes I, Prinz M, Pereira R et al (2006) Genetic analysis of three US population groups using a X-chromosomal STR decaplex. Int J Legal Med (in press)
10. Gusmão L, Sánchez-Diz P, Alves C, Lareu MV, Carracedo A, Amorim A (2000) Genetic diversity of nine STRs in two northwest Iberian populations: Galicia and northern Portugal. Int J Legal Med 114:109–113
11. Gusmão L, Butler J, Carracedo A et al (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int J Legal Med 120:191–200
12. Laan M, Wiebe V, Khusnutdinova E, Remm M, Pääbo S (2004) X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. Eur J Hum Genet 13:452–462
13. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
14. Lareu MV, Phillips CP, Carracedo A, Lincoln PJ, Syndercombe Court D, Thomson JA (1994) Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. Forensic Sci Int 66:41–52
15. Li WH, Yi S, Makova K (2002) Male-driven evolution. Curr Opin Genet Dev 12:650–656
16. Peloso G, Grignani P, Previdere C (2004) Allele distribution of five X-chromosome STR loci in an Italian population sample. Int Congr Ser 1261:260–262
17. Pepinski W, Skawronska M, Niemcunowicz-Janica A, Koc-Zorawska E, Janica J, Soltyszewski I (2005) Polymorphism of four X-chromosomal STRs in a Polish population sample. Forensic Sci Int 151:93–95
18. Pereira L, Prata MJ, Amorim A (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet 64:491–506
19. Pérez-Lezaun A, Calafell F, Clarimón J et al (2000) Allele frequencies of 13 short tandem repeats in population samples from the Iberian Peninsula and northern Africa. Int J Legal Med 113:208–214
20. Poetsch M, Sabule A, Petersmann H, Volksone V, Lignitz E (2006) Population data of 10 X-chromosomal loci in Latvia. Forensic Sci Int 157:206–209
21. Robino C, Giolitti A, Gino S, Torre C (2006) Development of two multiplex PCR systems for the analysis of 12 X-chromosomal STR loci in a northwestern Italian population sample. Int J Legal Med 120:315–318
22. Schaffner SF (2004) The X chromosome in population genetics. Nat Rev Genet 5:43–51
23. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462
24. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003a) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74
25. Szibor R, Plate I, Edelmann J, Hering S, Kuhlisch E, Michael M, Krause D (2003b) Chromosome X haplotyping in deficiency paternity testing principles and case report. Int Congr Ser 1239:815–820
26. Szibor R, Edelmann J, Zarrabeitia MT, Riancho JA (2003c) Sequence structure and population data of the X-linked markers DXS7423 and DXS8377—clarification of conflicting statements published by two working groups. Forensic Sci Int 134:72–73
27. Szibor R, Hering S, Kuhlisch E, Plate I, Demberger S, Krawczak M, Edelmann J (2005) Haplotyping of STR cluster DXS6801–DXS6809–DXS6789 on Xq21 provides a powerful tool for kinship testing. Int J Legal Med 119:363–369
28. Szibor R, Hering S, Edelmann J (2006) A new Web site compiling forensic chromosome X research is now online. Int J Legal Med 120:252–254
29. Turrina S, De Leo D (2003) Population data of three X-chromosomal STRs: DXS7132, DXS7133 and GATA172D05 in north Italy. J Forensic Sci 48:1428–1429
30. Vauhkonen H, Vauhkonen M, Sipponen P, Sajantila A (2004) Correlation between the allelic distribution of STRs in a Finnish

population and phenotypically different gastrointestinal tumours: a study using four X-chromosomal markers (DXS7423, DXS8377, ARA, DXS101). Ann Hum Genet 68:555–562
31. Wiegand P, Berger B, Edelmann J, Parson W (2003) Population genetic comparisons of three X-chromosomal STRs. Int J Legal Med 117:62–65
32. Zalán A, Völgyi A, Jung M, Peterman O, Pamjav H (2006) Hungarian population data of four X-linked markers: DXS8378, DXS7132, HPRTB, and DXS7423. Int J Legal Med (DOI 10.1007/s00414-005-0051-0)
33. Zarrabeitia M, Alonso A, Martin J et al (2006) Study of six X-linked tetranucleotide microsatellites: population data from five Spanish regions. Int J Legal Med 120:147–150
34. Zarrabeitia MT, Alonso A, Zarrabeitia A, Castro A, Fernandez I, Pancorbo MM (2004) X-linked microsatellites in two northern Spain populations. Forensic Sci Int 145:57–59