

## ARTICLE

# The X chromosome Alu insertions as a tool for human population genetics: data from European and African human groups

Georgios Athanasiadis<sup>1</sup>, Esther Esteban<sup>1</sup>, Marc Via<sup>1</sup>, Jean-Michel Dugoujon<sup>2</sup>, Nicholas Moschonas<sup>3</sup>, Hassen Chaabani<sup>4</sup> and Pedro Moral<sup>\*,1</sup>

<sup>1</sup>Unitat d'Antropologia, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain; <sup>2</sup>Centre d'Anthropologie, University Toulouse III, Toulouse, France; <sup>3</sup>Department of Biology, University of Crete, Herakleion, Greece; <sup>4</sup>Faculte de Pharmacie de Monastir, Universite du Centre, Monastir, Tunisia

Alu elements are the most abundant mobile elements in the human genome (~1 100 000 copies). Polymorphic Alu elements have been proved to be useful in studies of human origins and relationships owing to two important advantages: identity by descent and absence of the Alu element known to be the ancestral state. Alu variation in the X chromosome has been described previously in human populations but, as far as we know, these elements have not been used in population relationship studies. Here, we describe the allele frequencies of 13 'young' Alu elements of the X chromosome (Ya5DP62, Ya5DP57, Yb8DP49, Ya5a2DP1, Yb8DP2, Ya5DP3, Ya5NBC37, Yd3JX437, Ya5DP77, Ya5NBC491, Yb8NBC578, Ya5DP4 and Ya5DP13) in six human populations from sub-Saharan Africa (the Ivory Coast), North Africa (Moroccan High Atlas, Siwa oasis in Egypt, Tunisia), Greece (Crete Island) and Spain (Basque Country). Eight out of 13 Alu elements have shown remarkably high gene diversity values in all groups (average heterozygosities: 0.342 in the Ivory Coast, 0.250 in North Africa, 0.209 in Europe). Genetic relationships agree with a geographical pattern of differentiation among populations, with some peculiar features observed in North Africans. Crete Island and the Basque Country show the lowest genetic distance (0.0163) meanwhile Tunisia, in spite of its geographical location, lies far from the other two North African samples. The results of our work demonstrate that X chromosome Alu elements comprise a reliable set of genetic markers useful to describe human population relationships for fine-scale geographical studies. *European Journal of Human Genetics* (2007) 15, 578–583. doi:10.1038/sj.ejhg.5201797; published online 28 February 2007

**Keywords:** Alu elements; X chromosome; human populations; Mediterranean region

## Introduction

The X chromosome has the unique feature of being present only in a single copy in male subjects, which leads to a

series of special characteristics that justify its increasing interest in studies of human population genetics. Mutations occur less frequently in this chromosome because every existing X chromosome spends two-thirds of its lifetime in female subjects where the nucleotide mutation rate is much lower than in male subjects. Additionally, the effective population size of the X chromosome is three quarters that of the autosomes. These facts contribute to the lower diversity of this chromosome (estimated to be about half of that on the autosomes). However, the smaller

\*Correspondence: Professor P Moral, Department Biologia Animal-Antropologia, Faculty Biologia, Universidad de Barcelona, Avenue Diagonal 645, Barcelona 08028, Spain.

Tel: +34 934021461; Fax: +34 934035740;

E-mail: pmoral@ub.edu

Received 28 September 2006; revised 19 January 2007; accepted 24 January 2007; published online 28 February 2007

population size of the X chromosome also causes changes owing to genetic drift being faster than in other chromosomal regions and, therefore, population structure is expected to be more emphasized. Consequently, populations should differ more in their X chromosomes than in their autosomes. Linkage disequilibrium is also greater on the X chromosome, because only two-thirds of this chromosome manages to recombine in each generation. The size of regions with a single genetic history is expected to be larger than in autosomes, once more making it ideal for human population genetic studies. Finally, determining haplotypes on the X chromosome is a simple procedure and therefore it has proved to be a good choice for haplotype-based phylogenetic studies.<sup>1</sup>

X chromosome variation has been applied successfully in large-scale geographical studies that cope with the origin of non-African populations pointing in the direction of the 'Out of Africa' model.<sup>2,3</sup> However, the use of X chromosome for fine-scale geographical studies is negligible when compared with the high number of human population studies focused on Y chromosome and mtDNA variation.

Alu repeats are a category of short interspersed nuclear elements widely distributed in the genome of all primates. With a size of less than 500 bp, Alu elements are the most abundant mobile elements in the human genome (~1 100 000 copies). Twenty-five percent of the 'young' Alu human-specific elements have been incorporated in the human genome so recently that they are dimorphic for the presence or absence of the insertion. Subsequently, individuals can be polymorphic for the presence or absence of an Alu element at a particular chromosomal location. Alu elements have been shown to be useful in human evolution studies because they offer two important advantages compared with other polymorphic markers: (i) they are identical by descent – that is, individuals that share Alu-insertion polymorphisms have inherited them from a common ancestor and (ii) the ancestral state of each Alu insertion polymorphism is known to be the absence of the element, so that they can be used to 'polarize' the evolutionary process.<sup>4</sup>

The first thorough attempt to assess X chromosome variability relating to Alu elements was carried out in 2003 by Calliman *et al.*<sup>5</sup> The result was 345 Alu repeat elements from eight young Alu subfamilies, 264 of which were found on the X chromosome. From these elements, 16 were found to be polymorphic, with various levels of heterozygosity depending on the origin of each population. The heterozygosity data from that study suggested that the Alu elements from the X chromosome would be useful as genetic markers for human population genetics. The data showed a slight reduction in Alu polymorphism on the human sex chromosomes, in accordance with the observations made above. A brand new polymorphic Alu element, baptized DXS225 and belonging to the Ya5 Alu subfamily, was described by Pereira *et al.*<sup>6</sup> It was embedded in a LINE-1 retrotransposon, a region not previously examined, sug-

gesting that there might actually be more polymorphic Alu elements on the X chromosome.

So far, the potential of Alu polymorphisms of the X chromosome, including a wide number of Alu polymorphisms of several anthropologically well-defined human populations, has not been tested. The Mediterranean region appears to be an interesting sample on which to test the efficacy of these markers. Previous anthropological works on Alu polymorphisms seem to be quite controversial, with some authors defending genetic flow between various groups in the Mediterranean and others assuming less interpopulation genetic relationships.<sup>7–9</sup> In this region, there are two levels of genetic flow that could be the subject of interesting surveys: among Mediterranean groups, and between Mediterranean and sub-Saharan groups.

In this study, 13 Alu elements scattered along the whole X chromosome were analyzed for the first time in six anthropologically well-defined groups selected by geographical, historical and ethnical criteria, including the Basques, several Mediterranean groups and a sub-Saharan sample from the Ivory Coast. The objectives were: (i) to explore the variation of polymorphic X chromosome Alu repeats in well-defined human populations; (ii) to apply this variation to the study of population relationships among South European and North African groups, and the various sub-Saharan influences in these populations and (iii) to provide new evidence of the usefulness of these markers to uncover the genetic variation between human populations.

## Materials and methods

### DNA samples

Blood samples were collected for DNA extraction and Alu amplification from six different populations (525 individuals), each one originating from a different country. Samples were obtained, with informed consent, from healthy and unrelated participants of both sexes. All participants had all four of their grandparents born in the same region. The work was approved by the Ethical Committee of the University of Barcelona. Two of the populations originated from anthropologically well-defined Berber groups in Morocco and Egypt (High Atlas and Siwa Oasis, respectively). The other samples were from Monastir (a Centre-North region of Tunisia), the Basque Country (in northern Spain) and Crete Island (Greece). Finally, a sample from the Ahizi ethnic group from the Ivory Coast was genotyped in order to include a representation of the sub-Saharan African variation.

### Genetic determinations

PCR amplification was accomplished in 20- $\mu$ l reactions for the 13 Alu sequences of the X chromosome (Ya5DP62, Ya5DP57, Yb8DP49, Ya5a2DP1, Yb8DP2, Ya5DP3, Ya5NBC37, Yd3JX437, Ya5DP77, Ya5NBC491, Yb8NBC578, Ya5DP4 and Ya5DP13). Primer sequences are described in Calliman

et al,<sup>5</sup> as well as amplification and electrophoresis conditions with minor modifications.

### Statistical analysis

Allele frequencies were calculated by direct gene counting.  $\chi^2$  and Fisher's Exact tests were used to detect significant differences between male and female allele frequencies, and to check the Hardy–Weinberg equilibrium for each locus<sup>10</sup> using the GENEPOP v3.3<sup>11</sup> statistical package. Reynolds' distances<sup>12</sup> were calculated between pairs of populations with PHYLIP<sup>13</sup> statistical package and the consistency of the distance values was checked by bootstrap resampling analysis (100 iterations). Principal component analysis (PCA) was carried out using the R-MATRIX program.<sup>14</sup> Polymorphism and population structure within and between groups was tested by analysis of molecular variance (AMOVA) using the ARLEQUIN v2.0 program.<sup>15</sup> Global  $F_{ST}$  values were estimated by averaging partial values, and the resultant probability was calculated by combining probabilities from each individual test.<sup>16</sup>

Apart from the usual distance-based clustering methods described above, a model-based method was also used to infer population structure by means of the STRUCTURE 2.1 program.<sup>17</sup> A model of K population groups (where K might be unknown) was assumed. This model was tested for several values of K using a specific Markov Chain Monte Carlo algorithm (the Gibbs sampler). STRUCTURE estimates the 'natural logarithm of the probability of the data' for each value of K, briefly referred to as 'Ln P(X\K)'. Among the estimated K values, that yielding the lowest absolute value of the Ln P(X\K) is the one that best describes the data. In our data sets, we ran the Gibbs sampler under the admixture model (INFERALPHA = 1.0), using prior population information and assuming correlated allele frequencies. All runs included a burn-in period of 50 000 iterations followed by 10<sup>6</sup> iterations, and they were repeated three times each in order to test the consistency of the results.

### Results

#### Allele frequencies, Hardy–Weinberg equilibrium and heterozygosity

Table 1 shows the frequencies of the 13 Alu insertion alleles (Alu+) for the six populations. Some Alu elements appear to be fixed in several populations. In most cases, the frequency distributions fitted the Hardy–Weinberg equilibrium. The raw deviations found in Ya5DP62 in Ivory Coast ( $P=0.003$ ), Ya5DP57 in Basques ( $P=0.007$ ) and Ya5NBC491 in Tunisia ( $P=0.043$ ) were not significant after the Bonferroni correction. Heterozygosity levels per locus and per population are also shown in Table 1. In general, most Alu elements show moderate to high diversity, except for Ya5NBC491, Yb8NBC578, Ya5DP4 and Ya5DP13. The Ya5NBC37 Alu shows the highest

**Table 1** X chromosome Alu insertion frequencies and heterozygosity per locus and per population

	Ivory Coast	High Atlas	Siwa Oasis	Tunisia	Crete Island	Basque Country	H per locus (Ivory Coast)	H per locus (rest)
Ya5DP62	(82) 0.634	(147) 0.871	(123) 0.821	(168) 0.744	(116) 0.802	(87) 0.713	0.464	0.325
Ya5DP57	(84) 0.262	(140) 0.636	(137) 0.431	(159) 0.830	(86) 0.744	(115) 0.774	0.387	0.393
Yb8DP49	(85) 0.706	(149) 0.779	(138) 0.804	(168) 0.744	(120) 0.842	(136) 0.728	0.415	0.341
Ya5a2DP1	(85) 0.294	(151) 0.669	(140) 0.629	(159) 0.830	(119) 0.874	(136) 0.869	0.415	0.328
Yb8DP2	(83) 0.711	(151) 0.305	(142) 0.359	(165) 0.200	(121) 0.231	(138) 0.130	0.411	0.357
Ya5DP3	(87) 0.230	(151) 0.199	(142) 0.239	(164) 0.232	(120) 0.067	(134) 0.134	0.354	0.279
Ya5NBC37	(80) 0.313	(146) 0.432	(142) 0.275	(170) 0.412	(121) 0.339	(136) 0.324	0.430	0.452
Yd3IX437	(67) 0.552	(150) 0.213	(138) 0.246	(152) 0.033	(98) 0.143	(125) 0.104	0.495	0.240
Ya5DP77	(84) 0.286	(136) 0.912	(128) 0.914	(166) 0.892	(109) 0.982	(90) 1.000	0.400	0.093
Ya5NBC491	(67) 0.940	(148) 1.000	(143) 0.993	(158) 0.975	(121) 1.000	(132) 1.000	0.124	0.055
Yb8NBC578	(71) 0.789	(150) 0.940	(139) 0.950	(165) 0.933	(121) 1.000	(133) 0.932	0.333	0.092
Ya5DP4	(73) 0.000	(150) 0.013	(143) 0.000	(165) 0.012	(119) 0.059	(133) 0.067	0.000	0.047
Ya5DP13	(71) 0.901	(149) 0.993	(143) 1.000	(167) 0.988	(121) 1.000	(133) 1.000	0.195	0.009
H per population	0.3423 (0.1461)	0.2530 (0.1818)	0.2653 (0.1855)	0.2305 (0.1575)	0.2066 (0.1767)	0.2118 (0.1656)		

The parentheses before the frequencies show the number of chromosomes examined. The parentheses after the heterozygosities in the last row indicate the standard deviation. The heterozygosity per locus (two last columns) is indicated separately for the Ivory Coast and for the rest of the samples, owing to important heterozygosity differences between these two groups.

**Table 2** Reynolds' distances (below diagonal) and respective distance errors (above diagonal) for the six populations studied

	<i>Ivory Coast</i>	<i>Crete Island</i>	<i>Basque Country</i>	<i>Siwa Oasis</i>	<i>High Atlas</i>	<i>Tunisia</i>
Ivory Coast		0.0718	0.0768	0.0633	0.0564	0.0660
Crete Island	0.314490		0.0536	0.0255	0.0107	0.0089
Basque Country	0.322052	0.016299*		0.0265	0.0099	0.0055
Siwa Oasis	0.176117	0.072466	0.086411		0.0104	0.0347
High Atlas	0.215017	0.035995	0.048536	0.022322		0.0131
Tunisia	0.297528	0.029565	0.016804	0.086113	0.038674	

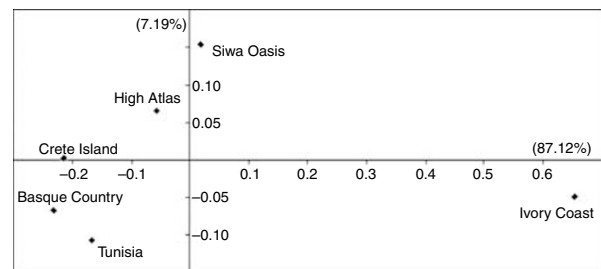
Only the distance between Basque Country and Crete Island was found not significantly different from zero marked with an asterisk.

heterozygosity (0.436), followed by Ya5DP57, Yb8DP49, Ya5a2DP1 and Yb8DP2 Alu elements. Likewise, the most diverse population seems to be the Ivory Coast ( $H=0.342$ ) followed by the rest of the North African (average  $H=0.250$ ) and South European populations (average  $H=0.209$ ). No significant linkage disequilibrium was present in any pair of the Alu markers, in accordance with the large chromosomal distances between them (ranging from 560 Kb to 59.65 Mb).

### Reynolds' distances and principal components analysis

Population pairwise comparisons are indicated in Table 2. Bootstrap resampling analysis for the calculation of distance errors showed high consistency of values. The lowest distance has been found between Crete Island and Basque Country (0.0163), and this is the only value not significantly different from zero. On average, the mean distance of Ivory Coast to the other five samples is 0.2650, whereas the mean distance among Mediterraneans and Basques is considerably lower (0.0453).

The PC analysis allowed the graphic representation of population relationships, as shown in Figure 1. The first two axes account for 94.31% of the total variance when the six samples are considered. The first axis clearly separates the sub-Saharan group of the Ivory Coast from the rest. The two European samples lie in the other extreme of the variation, whereas the three North African samples show an intermediate position, although always closer to the Europeans. The second component contributes to a relative separation among non-sub-Saharan groups, with Siwa Oasis and Tunisia appearing at the most distant positions. When the analysis was repeated, in order to remove the effect of the sub-Saharan sample (not shown), the relationship pattern among the remaining populations was substantially the same. In this case, the first axis (accounting for 64.79% of the total variance) underlines the separation of Siwa Oasis and High Atlas from the Basques, Crete Island and Tunisia. It is interesting to note that Tunisia is the North African group that lies genetically closest to the European samples.



**Figure 1** Principal components analysis of the Ivory Coast, Basque Country, Crete Island, High Atlas, Siwa Oasis and Tunisia. The first two axes account for 94.31% of the total variance. The percentage of variance accounted for by each axis is shown in parentheses.

### Analysis of molecular variance

A first approach to population diversity through the  $F_{ST}$  statistic yields a global value of 9.88% ( $P<0.001$ ) when all six populations were grouped together (data not shown). That is, almost 10% of the detected genetic variance was concentrated between populations. By locus, the  $F_{ST}$  values range from 0.99% ( $P<0.05$ , for Ya5NBC37) to an impressive 40.38% ( $P<0.0001$ , for Ya5DP77). The global  $F_{ST}$  (9.88) is mainly attributed to, on the one hand, the variation of Ya5DP57, Ya5a2DP1, Yb8DP2, Yd3JX437 and Ya5DP77 Alu markers, and, on the other hand, the inclusion of the Ivory Coast sample. By excluding the Ivory Coast, the global  $F_{ST}$  changed to a more moderate value (3.81%,  $P<0.001$ ). This mainly reflects the variation of the Ya5DP57 Alu polymorphism (individual  $F_{ST}=11.50\%$ ,  $P<0.0001$ ), which in Siwa Oasis presents clearly distinguishable allele frequencies.

A hierarchical AMOVA, assuming two geographical groups ('South European' and 'North African') without the Ivory Coast, does not reveal any significant differences between the two Mediterranean shores, the frequency variance between the two groups ( $F_{CT}=1.24\%$ ,  $P=0.338$ ) being clearly lower than the diversity among populations within groups ( $F_{SC}=3.06\%$ ,  $P<0.001$ ). Separate AMOVA tests for South European and North African samples gave  $F_{ST}$  values of 0.38 and 3.45, respectively, both of them significant ( $P<0.001$ ).

**Table 3** Estimated natural logarithm of the probability of the data ( $\ln P(X\backslash K)$ ) for each value of  $K$  in three data sets and proportion of membership of each pre-defined population in each of the three clusters

$K$	$\ln P(X\backslash K)$				
	All populations	All Mediterranean populations	Cluster 1	Cluster 2	Cluster 3
1	-4178.6	-3404.0			
2	-4043.4	-3531.9			
3	-4007.2	-3528.4			
4	-4094.6	-3723.2			
5	-4317.1	-3937.1			
6	-4450.8	—			
Ivory Coast	0.770	0.126	0.104		
Crete Island	0.194	0.379	0.427		
Basque Country	0.188	0.422	0.390		
Tunisia	0.208	0.397	0.394		
Siwa Oasis	0.388	0.300	0.312		
High Atlas	0.323	0.324	0.353		

### Model-based inference of population structure

To explore the degree of genetic structure among our samples, we defined a STRUCTURE data file containing all six populations. We estimated the posterior probabilities departing from the  $\ln P(X\backslash K)$  value. As seen in Table 3, when surveying population structure in all six populations, the model with  $K=3$  seems to fit our data best. The same table also indicates the proportion of membership of each pre-defined population in each of the three clusters inferred for data set. The pattern of membership of the Ivory Coast is the most differentiated, with 77% of membership in cluster 1. The European and Tunisian samples show a similar pattern of membership in the three clusters (1:2:2), whereas Siwa Oasis and High Atlas seem to follow a 1:1:1 pattern. As expected, no cluster gets to become exclusively characteristic of the populations implicated, although cluster 1 could be considered as representative of the sub-Saharan variation. To delve into the two clusters not directly related with the sub-Saharan variation, we repeated the runs including only the five non-sub-Saharan groups (see Table 3). This time the model with  $K=1$  was best to describe the data.

### Discussion

This paper describes the pattern of the frequency distribution of 13 polymorphic Alu insertions of the X chromosome in five well-defined groups from the Mediterranean region, including the Basque Country. In general, the allele frequencies found range within the general patterns described previously,<sup>5</sup> but with a remarkable between-population variation. In terms of variation, the six samples were adjusted to a decreasing pattern of diversity from

South to North (mean heterozygosity for the Ivory Coast: 0.342, for North African samples: 0.250 and for the European ones: 0.209).

The hierarchical AMOVA analysis in the five Mediterranean groups (Basques included) showed that only a small and nonsignificant part of the genetic variance could be attributed to the variation between North–South groups ( $F_{CT}=1.24\%$ ), indicating no particular genetic differentiation between the two sides of the Mediterranean Sea. However, the markers examined are consistent with a more important diversity within North Africa ( $F_{ST}=3.45\%$ ,  $P<0.001$ ) than in South Europe (0.38%,  $P<0.001$ ). Comas *et al*<sup>7</sup> and González-Pérez *et al*<sup>8</sup> had previously studied population relationships in the western-Mediterranean basin using polymorphic autosomal Alu elements. These two surveys indicated a North *versus* South differentiation ( $F_{CT}$  values: 1.80 and 1.96%, respectively) slightly higher than in our results, but a population variation within groups clearly lower ( $F_{SC}$  values of 2.30 and 0.47%, respectively) than that evidenced from the X chromosome Alu markers ( $F_{SC}$  of 3.06%). Moreover, the 13 Alu markers of this study reveal a population variation within North Africa ( $F_{ST}=3.45\%$ ) sixfold higher than that obtained from a similar number of autosomal Alu elements ( $F_{ST}=0.57\%$ ; Comas *et al*, 2000). The X chromosome markers reveal a higher population differentiation in comparison with the same kind of genetic markers in autosomes. This possibly reflects the effect of the reduced population size of the X chromosome on population variation, which, as we mentioned in the introduction, makes populations differ more in their X chromosomes than in autosomal markers.

The Reynolds' distances revealed generally accepted relationships among the Mediterranean populations. It is remarkable that the Arab-speaking sample from Tunisia shows a particular genetic position as compared with other North African groups. In fact, the Tunisian genetic distances to European samples are smaller than those to North African groups. This close position of Tunisia to the Europeans also appears in the population distribution in the PCA graph (Figure 1). This could be explained by the history of the Tunisian population, reflecting the influence of the ancient Phoenician settlers of Carthage followed, among others, by Roman, Byzantine, Arab and French occupations, according to historical records. Notwithstanding, other explanations cannot be discarded, such as the relative heterogeneity within current Tunisian populations,<sup>18</sup> and/or the limited sub-Saharan genetic influence in this region as compared with other North African areas, without excluding the possibility of the genetic drift, whose effect might be particularly amplified on the X chromosome.

An interesting aspect comes from the evidenced relationships between the Basque Country and Crete Island. These two populations have distinct historical, anthropological and cultural backgrounds, and yet no significant differ-

ences were found between them when a locus-by-locus  $\chi^2$  comparison was carried out. As for the remaining analyzed populations, Siwa Oasis seems to be the most differentiated (see Table 2 and Figure 1). The differentiation shown by Siwa Oasis, and also by High Atlas, could be related to higher foreign genetic contributions, from West Sahara into High Atlas and Nile groups into the Siwa Oasis. Esteban *et al*<sup>19</sup> described a similar pattern of GGC allele frequencies of the androgen receptor (located in chromosome X) for the Ivory Coast and Siwa Oasis samples, giving evidence of sub-Saharan genetic influence in this Berber group.

The model-based method indicated population structure with three clusters inferred when all populations were examined. Cluster 1 (Table 3) is evidently the 'sub-Saharan' one. As for the other two, they seem to have a biological meaning only when seen in comparison with cluster 1. All European and North African samples show a 1:1 membership proportion in clusters 2 and 3. Only when seen in conjunction with cluster 1 do differences appear. The Basque country, Crete Island and Tunisia feature a similar pattern of membership in the three clusters (1:2:2), whereas Siwa Oasis and High Atlas seem to follow a 1:1:1 pattern. This could possibly explain why the software has failed to detect population structure in the Mediterranean populations, implying that there are no population-specific genetic patterns representative enough to allow us to assign, with certainty, individuals to populations.

The failure to detect population structure among Mediterranean groups might lead to the conclusion that our data are controversial, because we have already discussed the striking differences of, for example, Siwa Berbers and High Atlas Berbers from the other groups. Apparently, differences do exist, but they are not striking enough to allow the definition of different clusters within the Mediterranean region. Clustering appears only when a quite distinct human group is added, such as the sub-Saharan Ahizi from the Ivory Coast.

To sum up, our data on X chromosome markers support, in general, the differentiation patterns of the Mediterranean populations described by other investigators, providing, at the same time, detailed data of the frequency distribution of X chromosome Alu elements. To our knowledge, it is the first time that these specific molecular markers have been used in such a study. X chromosome Alu elements seem to perform well in fine-scale population differentiation studies. As no completely comparable data exist to survey the effect of the reduced population size of the X chromosome on the genetic distances between populations, a future investigatory line should include data from autosomal markers from the same populations. Furthermore, linkage disequilibrium studies are advisable, including higher mutation rate STR markers close to the X chromosome Alu insertions.

## Acknowledgements

We are grateful to all of the donors for providing blood samples and the people who contributed to the collection. In particular, we thank Professor André Chaventré and Dr Gil Bellis (for the samples from Ivory Coast). This research has been supported by the Ministerio de Ciencia y Tecnología CGL 2005-3391 and Generalitat de Catalunya SGR00252 projects. The sampling of the Berbers from Morocco and Egypt was supported by the Conseil Régional de Midi-Pyrénées, Toulouse (France). The work of GA has been financed by an FPU grant from the Ministerio de Educación y Ciencia (grant reference: AP2005-4425).

## References

- Schaffner SF: The X chromosome in population genetics. *Nat Rev Genet* 2004; 5: 43–51.
- Harris EE, Hey J: X chromosome evidence for ancient human histories. *Proc Natl Acad Sci U S A* 1999; 96: 3320–3324.
- Yu N, Fu YX, Li WH: DNA polymorphism in a worldwide sample of human X chromosomes. *Mol Biol Evol* 2002; 19: 2131–2141.
- Batzer MA, Deininger PL: Alu repeats and human genomic diversity. *Nat Rev Genet* 2002; 3: 370–379.
- Calliman PA, Hedges DJ, Salem AH *et al*: Comprehensive analysis of Alu-associated diversity on the human sex chromosomes. *Gene* 2003; 317: 103–110.
- Pereira RW, Santos SS, Pena SD *et al*: A novel polymorphic Alu insertion embedded in a LINE 1 retrotransposon in the human X chromosome DXS225: identification and worldwide population study. *Genet Mol Res* 2006; 5: 63–71.
- Comas D, Calafell F, Benchemi N *et al*: Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 2000; 107: 312–319.
- Gonzalez-Perez E, Via M, Esteban E *et al*: Alu insertions in the Iberian Peninsula and north west Africa – genetic boundaries or melting pot? *Coll Antropol* 2003; 27: 491–500.
- Garcia-Obregon S, Alfonso-Sanchez MA, Perez-Miranda AM *et al*: Genetic position of Valencia (Spain) in the Mediterranean basin according to Alu insertions. *Am J Hum Biol* 2006; 18: 187–195.
- Guo SW, Thompson EA: Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 1992; 48: 361–372.
- Raymond M, Rousset F: GENEPOP version 1.2 population genetics software for exact tests and ecumenicism. *J Hered* 1995; 86: 248–249.
- Reynolds J, Weir BS, Cockerman CC: Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* 1983; 105: 767–779.
- Felsenstein J: PHYLIP – phylogeny inference package. *Cladistics* 1989; 5: 164–166.
- Harpending H, Jenkins T: Genetic distance among southern African populations; in Crawford MH, Workman PL (eds): *Methods and Theories of Anthropological Genetics*. University of New Mexico Press, 1973, pp 177–199.
- Schneider S, Roessli D, Excoffier L: Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Geneva, 2000.
- Sokal RR, Rohlf FJ: *Biometry*. Second edition, Freeman and Co, New York, 1981.
- Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000; 155: 945–959.
- Giraldo MP, Esteban E, Aluja MP *et al*: Gm and Km alleles in two Spanish Pyrenean populations (Andorra and Pallars Sobira): a review of Gm variation in the Western Mediterranean basin. *Ann Hum Genet* 2001; 65: 537–548.
- Esteban E, Rodon N, Via M *et al*: Androgen receptor CAG and GGC polymorphisms in Mediterraneans: repeat dynamics and population relationships. *J Hum Genet* 2006; 51: 129–136.