Hum Genet (2000) 107 : 312–319
DOI 10.1007/s004390000370

ORIGINAL INVESTIGATION

**David Comas · Francesc Calafell · Noufissa Benchemsi
Ahmed Helal · Gerard Lefranc · Mark Stoneking
Mark A. Batzer · Jaume Bertranpetit · Antti Sajantila**

Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits

Received: 9 May 2000 / Accepted: 18 July 2000 / Published online: 15 September 2000

Abstract An analysis of 11 Alu insertion polymorphisms (ACE, TPA25, PV92, APO, FXIIIB, D1, A25, B65, HS2.43, HS3.23, and HS4.65) has been performed in several NW African (Northern, Western, and Southeastern Moroccans; Saharawi; Algerians; Tunisians) and Iberian (Basques, Catalans, and Andalusians) populations. Genetic distances and principal component analyses show a clear differentiation of NW African and Iberian groups of samples, suggesting a strong genetic barrier matching the geographical Mediterranean Sea barrier. The restriction to gene flow may be attributed to the navigational hazards across the Straits, but cultural factors must also have played a role. Some degree of gene flow from sub-Saharan Africa can be detected in the southern part of North Africa and in Saharawi and Southeastern Moroccans, as a result of a continuous gene flow across the Sahara desert that has created a south-north cline of sub-Saharan Africa influence in North Africa. Iberian samples show a substantial degree of homogeneity and fall within the cluster of European-based genetic diversity.

D. Comas · F. Calafell · J. Bertranpetit (✉)
Unitat de Biologia Evolutiva.
Facultat de Ciències de la Salut i de la Vida.
Universitat Pompeu Fabra. Doctor Aiguader 80.
08003 Barcelona, Catalonia, Spain
e-mail: jaume.bertranpetit@cexs.upf.es,
Tel.: +34 93 542 2840, Fax: +34 93 542 2802

D. Comas · A. Sajantila
Department of Forensic Medicine, P.O. Box 40,
University of Helsinki, 00014 Helsinki, Finland

N. Benchemsi
Centre National de Transfusion Sanguine, 11400 Rabat, Morocco

A. Helal
Faculté de Pharmacie de Monastir. 5019 Monastir. Tunisia

G. Lefranc
CNRS UPR 1142 – Institut de Génétique Humaine
et Université Montpellier II, 34095 Montpellier Cedex 5, France

M. Stoneking
Max Planck Institute for Evolutionary Anthropology,
Inselstrasse 22, D-04103 Leipzig, Germany

M. A. Batzer
Departments of Pathology, Biometry and Genetics,
Biochemistry and Molecular Biology,
Stanley S. Scott Cancer Center,
Neuroscience Center of Excellence,
Louisiana State University Health Sciences Center
1901 Perdido Street, New Orleans, LA 70112, USA

Introduction

The population history of North Africa is particularly interesting because, although the region belongs to continental Africa, its history has been completely different from the sub-Saharan part. The peopling of the region has been influenced by two strong geographical barriers: the Sahara Desert to the south, which splits the African continent into two differentiated regions, and the Mediterranean Sea to the north, which separates the European and African continents. These geographical barriers may have constrained human movements in North Africa into an east-west gradient, although they were not impermeable to human movements. During the first half of the Holocene, the humid climate that prevailed in the Sahara produced a receding of the desert allowing human settlements, but over the past 5000 years, the Sahara Desert has suffered a gradual aridification and has become as dry as it is nowadays (Said and Faure 1990). Historical records document extensive trade routes that were established across the desert between sub-Saharan Africa and the north coast. In contrast, since the time of the Phoenicians, the city-based settlement pattern of the NW African coast integrated the area into the Mediterranean world. The seaward orientation of populations persisted and, similar to the desert, separated the Maghreb (NW Africa) from the rest of Africa to the south (Newman 1995). Moreover, during the 8th century AD, Berbers from North Morocco and Algeria under Arab leadership crossed the Mediterranean Sea and occupied the Iberian Peninsula for almost eight centuries, although the demographic impact of the conquest is thought to be limited (Hitti 1990).

Until recently, few genetic studies have been performed in NW Africa. In the latest compilation of classi-

cal genetic markers in North Africa (Bosch et al. 1997), the first principal component (PC) of gene frequencies showed an east-west pattern of genetic differentiation, in agreement with the geographical barrier imposed by the Sahara and the Mediterranean. Recent work with autosomal short tandem repeats (STRs; Bosch et al. 2000), mitochondrial DNA (mtDNA) sequences (Rando et al. 1998), and Y-chromosome haplotypes (Bosch et al. 1999) has suggested that the gene flow between NW Africa and Iberia and that between sub-Saharan Africa and NW Africa has been small. MtDNA variation in NW Africa (Rando et al. 1998) has shown a high frequency (up to 25%) of geographically specific sequences (named haplogroup U6) that is essentially absent in the Iberian Peninsula (from 0% in Andalusians to 5% in Portuguese). The mtDNA analysis has shown a limited gene flow from Europe to NW Africa that could be attributed to recent human movements. The study of Y-chromosome haplotypes (Bosch et al. 1999) shows little admixture between NW Africa and the Iberian Peninsula. The study of 21 autosomal STR loci in NW Africa has also shown a clear genetic difference between NW African populations and Iberians, although some degree of gene flow into Southern Iberia (Andalusians) can be detected (Bosch et al. 2000).

There are a large number of Alu insertion polymorphisms throughout the human genome; these are rapid and easy to type, apparently selectively neutral, and have known ancestral states. The insertion of an Alu element into the human genome is almost certainly a unique event, making any pair of Alu insertion alleles identical by descent and free of homoplasy (Batzer and Deininger 1991; Batzer et al. 1994; Stoneking et al. 1997). The use of these polymorphisms in a world-wide survey of human populations has confirmed the African origin of modern humans (Batzer et al. 1994, 1996; Stoneking et al. 1997). However, the use of Alu insertion polymorphisms in human evolution has been focused world-wide, and except for some population studies (Novick et al. 1998), relatively little research has been devoted to specific population questions. We have analyzed several NW African and Iberian populations for 11 Alu insertion polymorphisms, three of which have not been analyzed in previous world-wide population studies, in order: (1) to determine the genetic differentiation of the Alu insertion polymorphisms in NW African populations; (2) to compare the genetic composition of NW African and Iberian populations, establishing the possible amount of gene flow between them; and (3) to detect possible admixture from sub-Saharan African populations into NW Africa. The use of polymerase chain reaction (PCR)-based neutral autosomal DNA polymorphisms whose frequencies are only dependent on drift and migration (and not mutation) and whose ancestral state is known is a novel and powerful tool for the study of human populations.

Materials and methods

A total of 676 autochthonous individuals from NW Africa and the Iberian Peninsula was analyzed (Fig. 1). The sample comprised unrelated healthy blood donors, and informed consent was obtained from all individuals participating in the study. The Iberian samples comprised 96 Basques, 60 Catalans, and 67 Andalusians. The Moroccan samples were divided into three different groups according to their origin: 140 individuals from the western part of the country, a region mostly inhabited by Arabs, 111 individuals from the Rif mountains, in the north of Morocco, and 49 individuals from the Atlas mountain range, in the southeastern part of Morocco, both areas being inhabited mostly by Berber people, and 58 individuals from Western Sahara. Samples from Tunisia (48 individuals) and Algeria (47 individuals) were also analyzed. Additional African and European groups were as described by Stoneking et al. (1997).

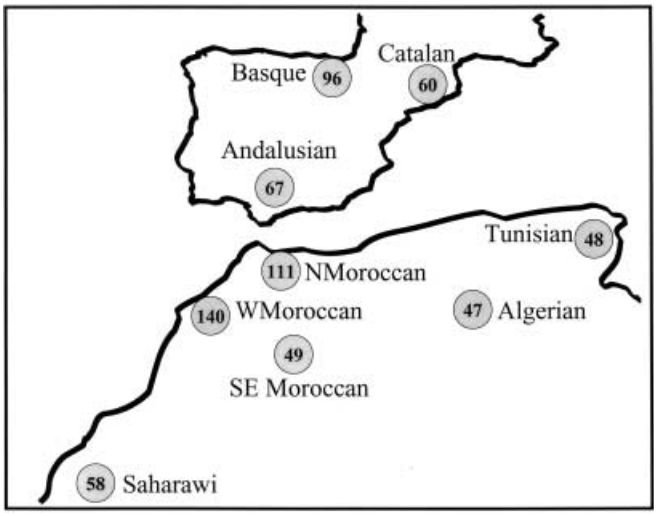

Fig. 1 Location of the samples typed for the 11 Alu insertion loci. *Numbers in circles* Sample sizes

Eleven human-specific Alu insertion polymorphisms (A25, B65, ACE, D1, APO, FXIIIB, PV92, TPA25, HS2.43, HS3.23, and HS4.65) were typed in each sample by using the primers described previously (Arcot et al. 1995a, 1995b, 1996; Batzer et al. 1996). The PCR amplification conditions for the first eight loci were performed as described previously (Stoneking et al. 1997) and for the other three loci (HS2.43, HS3.23, and HS4.65) were as follows: 95°C for 2 min, 52°C for 1 min, and 72°C for 1 min during 30 cycles, with a final elongation step of 72°C for 7 min.

Allele frequencies were calculated by direct counting; the Hardy-Weinberg equilibrium was assessed by an exact test (Guo and Thompson 1992) provided by the Arlequin program (Schneider et al. 1996). Gene diversity for each population and each locus was calculated according to the formula $D=n/(n\text{-}1)\ (1\text{-}\Sigma\, x^2)$, where n is the number of gene copies in the sample and x the frequency of each allele.

F_{ST}-related genetic distances were computed between pairs of populations (Reynolds et al. 1983) and were represented in a neighbor-joining (NJ) tree (Saitou and Nei 1987) by means of the PHYLIP 3.5c package (Felsenstein 1989). The tree topology was assessed through 1000 bootstrap iterations. Principal component analysis (PCA) was performed on the correlation matrix of the Alu insertion frequencies analyzed by using the SPSS package.

In order to ascertain the proportion of the genetic variance attributable to differences within or between populations, genetic variance was hierarchically apportioned through the analysis of molecular variance (AMOVA; Excoffier et al. 1992) performed with the Arlequin program (Schneider et al. 1996).

A Delaunay network was built in order to identify the zones of sharpest genetic change (see Bosch et al. 1997). To construct the network, we defined pairs of contiguous samples and connected them by a total of 15 edges. The genetic distance between each pair of samples was assigned to each edge, and the sharpest genetic

Table 1 Alu insertion frequencies in nine populations of the Iberian Peninsula and NW Africa (*2N* sample size in number of chromosomes typed)

Population	2N	ACE	TPA25	PV92	APO	FXIIIB	D1	A25	B65	HS2.43	HS3.23	HS4.65
Basques	192	0.443	0.568	0.188	0.953	0.484	0.380	0.219	0.604	0.125	0.875	0.005
Catalans	120	0.300	0.608	0.175	0.983	0.500	0.350	0.125	0.525	0.083	0.867	0.058
Andalusians	134	0.470	0.590	0.194	0.985	0.448	0.306	0.142	0.552	0.067	0.858	0.037
Northern Moroccans	222	0.333	0.617	0.333	0.910	0.338	0.288	0.113	0.608	0.045	0.833	0.081
Western Moroccans	280	0.314	0.575	0.343	0.929	0.293	0.304	0.143	0.614	0.071	0.821	0.114
Southeastern Moroccans	98	0.265	0.510	0.398	0.847	0.306	0.194	0.235	0.510	0.020	0.878	0.122
Saharawi	116	0.284	0.397	0.310	0.836	0.371	0.259	0.138	0.534	0.009	0.862	0.112
Algerians	94	0.266	0.532	0.287	0.915	0.315	0.149	0.106	0.734	0.085	0.840	0.032
Tunisians	96	0.240	0.604	0.313	0.875	0.344	0.245	0.167	0.594	0.052	0.750	0.063

boundary was obtained by tracing a perpendicular line across the edges showing the highest genetic distance. The procedure was iterated to calculate the second and the third genetic boundaries.

Results

Frequencies of the 11 human-specific Alu insertion polymorphisms in the nine populations typed (Fig. 1) are shown in Table 1. All loci are biallelic, and only the frequency of the presence of the Alu insertion is shown. All loci were polymorphic in all populations. Eight out of 99 tests for Hardy-Weinberg equilibrium showed significant departures from equilibrium. After application of the Bonferroni correction, only two comparisons gave significant departures (D1 in Basques and A25 in Saharawi). As none of the departures cluster by locus or by population, they probably represent random statistical fluctuations. Table 2 shows the average gene diversity by locus and population. Gene diversities in a biallelic locus are a direct function of the frequency of any of the two alleles. The loci analyzed in the present samples show significant differences in their gene diversity (Kruskal-Wallis test, $P<0.001$), which is a consequence of the observation that, at some loci, both alleles have similar frequencies, whereas in others, one of the alleles is rarer, because of random fluctuations. Nevertheless, when we focus on average gene diversity by population, no significant differences between samples are found (Kruskal-Wallis' test, $P=0.999$), because of the similar Alu insertion frequencies found in all the samples analyzed.

In order to asses the relationship between the populations analyzed, F_{ST}-genetic distances were calculated (data not shown) and depicted in an NJ tree (Fig. 2). The tree clearly divides the populations into two groups: North African populations and Iberians. The node separating the groups shows a strong bootstrap support after 1000 iterations (98.7%). Aside from this node, no other nodes within the two groups show strong bootstrap support, except for that linking Saharawi and Southeastern Moroccans to the rest of the network.

Table 2 Average gene diversity by locus and population

Locus		Population	
ACE	0.265±0.020	Basques	0.332±0.054
TPA25	0.493±0.003	Catalans	0.312±0.053
PV92	0.399±0.025	Andalusians	0.316±0.056
APO	0.154±0.030	Northern Moroccans	0.330±0.046
FXIIIB	0.467±0.011	Western Moroccans	0.338±0.042
D1	0.398±0.023	Southeastern Moroccans	0.344±0.045
A25	0.265±0.020	Saharawi	0.341±0.045
B65	0.485±0.010	Algerians	0.337±0.042
HS2.43	0.114±0.020	Tunisians	0.301±0.045
HS3.23	0.270±0.017		
HS4.65	0.125±0.024		

Since an NJ tree imposes a bifurcating model onto a distance matrix, which may be inadequate for closely related populations, we also assessed the genetic relationship among the populations through a PCA. The first two PCs account for 68.8% of the genetic variance observed, and their plot (Fig. 3) shows a similar pattern to that displayed in the NJ tree. The first PC clearly separates the Iberian populations that are characterized (with an absolute correlation greater than 0.85) by high frequencies of APO and FXIIIB Alu insertions, and low PV92 frequencies. The singularity of Saharawi and Southeastern Moroccans with respect to the rest of NW African populations, shown by the second PC, appears to be (with $r=-0.87$) attributable to the low frequency of the B65 Alu insertion.

As expected from the previous results, the first genetic boundary in the Delaunay network (Fig. 4) separates Iberian and North African samples. The second separates the Saharawi and the Southeastern Moroccans from the rest, and the third separates North African populations in an east-west split.

An AMOVA was performed considering all the samples as a single group in order to establish the apportionment of the genetic variance. The fraction of the genetic variance resulting from differences between populations

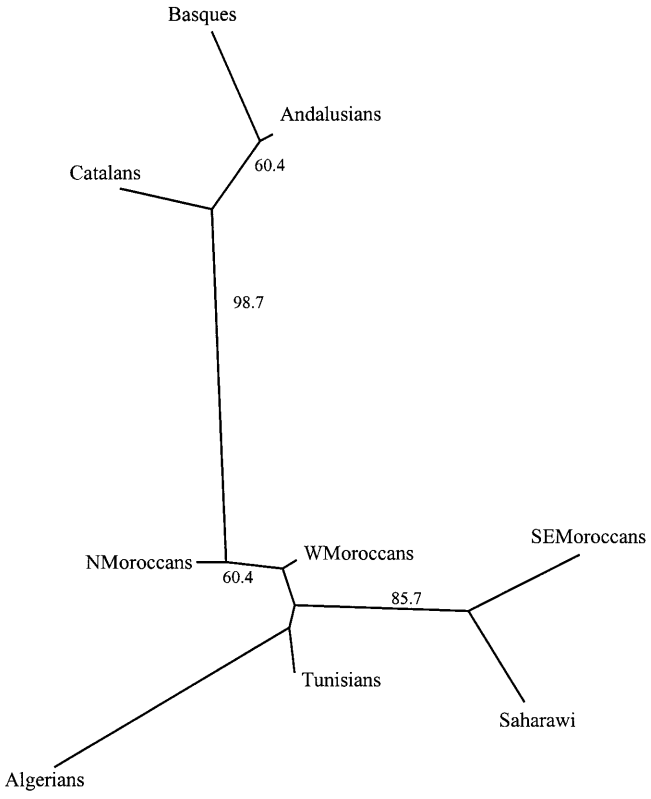

Fig. 2 Neighbour-joining tree of the genetic distance matrix for North African and Iberian populations for the 11 Alu insertion loci analyzed (*NMoroccans* Northern Moroccans, *WMoroccans* Western Moroccans, *SEMoroccans* Southeastern Moroccans). Bootstrap supports over 60% (out of 1000 iterations) are indicated along the *nodes*

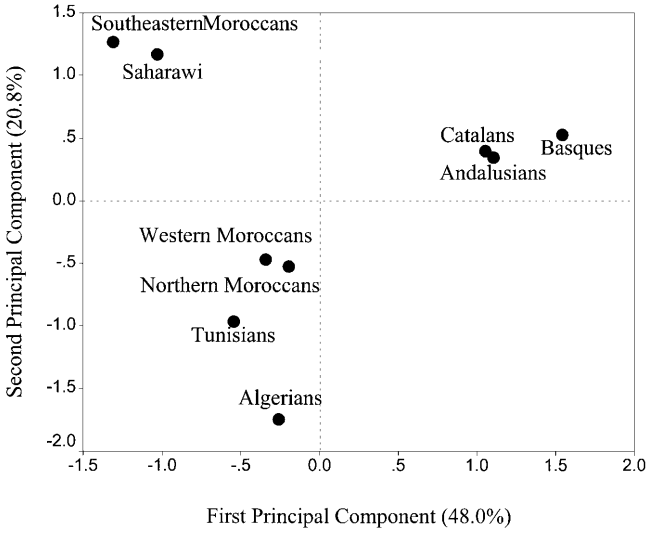

Fig. 3 Plot of the first two principal components of the allele frequencies at the 11 Alu insertion loci in several North African and Iberian populations

was 1.48% (a value significantly different from zero, $P<0.001$), whereas the rest was found within populations. When the populations were divided into Iberians and North Africans, the fraction of the genetic variance attributable to differences among groups was 1.96% ($P=0.009$), whereas differences among populations within groups was 0.47% ($P<0.001$). If we restricted the AMOVA to the Iberian populations, the genetic variance attributable to differences among populations was not significantly different from zero (0.26%, $P=0.172$), suggesting that Iberian populations are highly homogeneous. Within North Africa, 0.57% of the genetic variance was attributable to differences among populations ($P<0.001$).

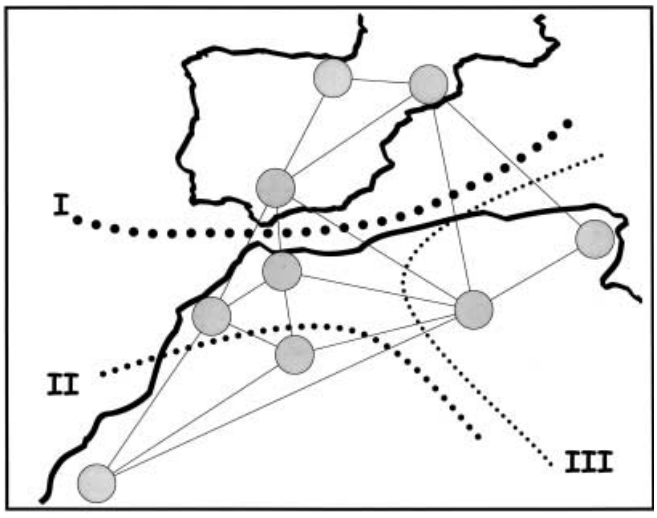

Fig. 4 Delaunay triangulation between the geographical localities for the samples analyzed. The first (*I*), second (*II*), and third (*III*) most significant genetic boundaries recognized on the basis of genetic distances are shown as *dashed lines*

In order to place NW African and Iberian Alu insertion genetic diversity in a global frame, we compared our samples with other European and African groups: French, Bretons, and Swiss, and !Kung, Nguni, and Bantu speakers from the Sotho/Tswana branch. The Alu insertion frequencies for the eight loci analyzed by Stoneking et al. (1997) were compiled, and the three remaining loci (HS2.43, HS 3.23, and HS4.65) were typed as described with the same samples. The Alu insertion frequencies for the six samples used for comparison are shown in Table 3. Genetic distances were calculated and represented in an NJ tree (Fig. 5). The tree clearly separated first sub-Saharan populations from the rest with strong bootstrap support (96.9%) and subsequently Europeans from NW Africans (85.8% of bootstrap support). The Iberian samples clustered together with the rest of European populations, and no robust branches existed between them, suggesting a high homogeneity within European populations. NW African populations clustered together; however, it is also interesting to note that the two samples that presented unique characteristics in the previous analyses (Saharawi and Southeastern Moroccans) were closer to sub-Saharan Africans, and the branch linking them together showed strong bootstrap support (82.9%). This affinity suggested a certain degree of gene flow from sub-Saharan Africa into NW Africa. When a hypothetical ancestral popula-

Table 3 Alu insertion frequencies for the loci HS2.43, HS3.23, and HS4.65 for six sub-Saharan African and European samples used for comparison. Sample size in number of chromosomes typed are shown in *brackets*

	HS2.43	HS3.23	HS4.65
French	0.092 (142)	0.894 (142)	0.014 (144)
Bretons	0.097 (144)	0.887 (142)	0.007 (144)
Swiss	0.022 (138)	0.831 (136)	0.028 (142)
!Kung	0.000 (84)	0.950 (80)	0.066 (76)
Nguni	0.000 (70)	0.912 (34)	0.000 (44)
Bantu speakers	0.000 (96)	0.945 (94)	0.182 (88)

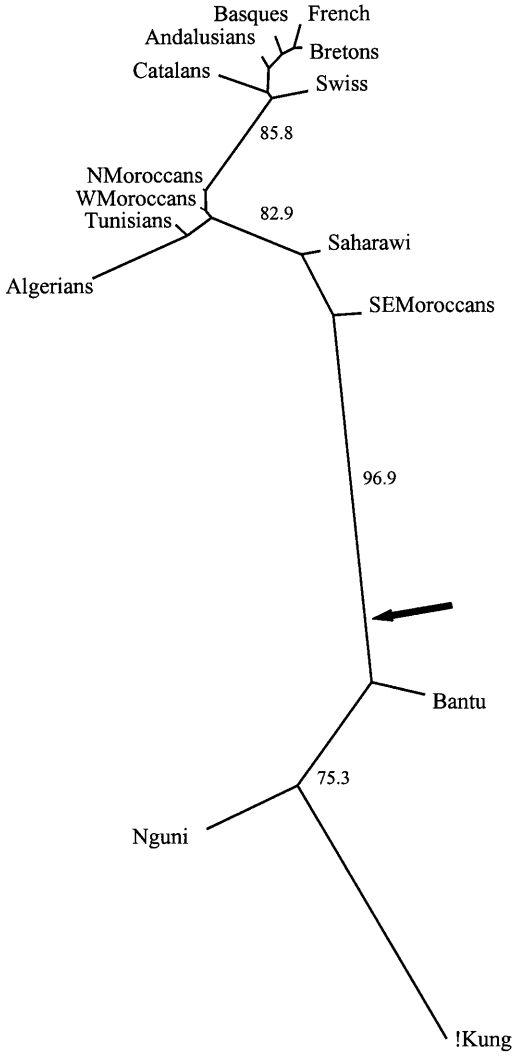

Fig. 5 Neighbour-joining tree of the genetic distance matrix for European and African populations for the 11 Alu insertion loci analyzed (*NMoroccans* Northern Moroccans, *WMoroccans* Western Moroccans, *SEMoroccans* Southeastern Moroccans). Bootstrap supports over 60% (out of 1000 iterations) are indicated along the *nodes*. *Arrow* Position of the hypothetical ancestor population that does not contain Alu insertions

tion that did not contain any of the Alu repeats was added, the resulting branch fell between sub-Saharan African and the rest of populations (position denoted by an arrow in Fig. 5), as described in previous studies (Batzer et al. 1994, 1996; Stoneking et al. 1997).

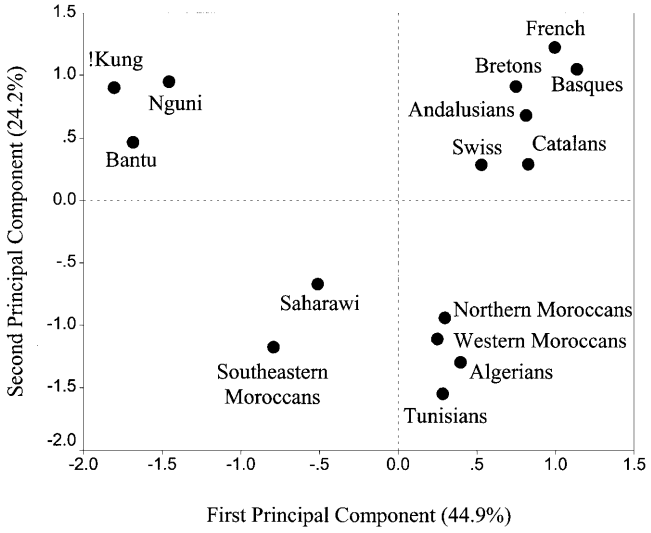

Fig. 6 Plot of the first two principal components of the allele frequencies at the 11 Alu insertion loci in several African and European populations

The plot of the first two PCs, which encompassed 69.1% of the genetic variance observed, showed a similar display (Fig. 6). The first PC separated sub-Saharan African populations from the rest of the samples, but Saharawi and Southeastern Moroccans were located in an intermediate position. Sub-Saharan samples were characterized (with absolute correlations greater than 0.8) by high frequencies of Alu insertions at locus A25 and low frequencies at APO, FXIIIB, HS2.43, and TPA25. The second PC encompassed 24.2% of the observed variance and separated NW African samples, which were characterized (with absolute correlations greater than 0.7) by high frequencies of the PV92 Alu insertion and low frequencies of the ACE insertion. When a hypothetical ancestral population was added to the PC analysis (data not shown), the plot of the first two PCs showed a similar display, but in this case, the first PC separated the ancestral from the rest of the populations, and the second PC separated sub-Saharan Africans from the rest.

Discussion

In this paper, we have analyzed 11 Alu insertion polymorphisms in several North African and Iberian populations, and in all the analyses performed, a clear differentiation between both groups of populations has been detected. These results are concordant with previous studies of classical polymorphisms (Bosch et al. 1997; Simoni et al. 1999), Y-chromosome polymorphisms (Bosch et al. 1999), and autosomal STRs (Bosch et al. 2000): a clear differentiation between NW Africa and Iberians is shown by genetic distances and PC analysis. Although a deep NW African cultural influence in the Iberian Peninsula is well

documented by historians, its demographic impact may have been limited. The geographical distance between NW Africa and the Iberian Peninsula at its narrowest part is less than 15 km, but it could have acted as a strong geographical barrier, hindering gene flow between the two continents. The Mediterranean is a closed sea with high evaporation and draws water from the Atlantic Ocean through the Gibraltar Straits. This fact produces a strong maritime current that might have made navigation difficult and restricted gene flow. Even if sailing difficulties are unquestioned, the distance is short enough to allow ample migration. The present analysis, however, shows that the Gibraltar Straits acted as a genetic discontinuity. Nonetheless, it does not imply that the geographic barrier by itself prevented migration. A genetic barrier by itself may not genetically differentiate the populations that it divides. A different mechanism may have generated the genetic differentiation across the Gibraltar Straits: the Neolithic wave of advance may have run in parallel along the two Mediterranean shores (Bosch et al. 1997; Simoni et al. 1999). This phenomenon may also have generated a cultural difference by bringing Indoeuropean languages to the Northern Mediterranean shore and Afroasiatic languages to the Southern shore (Renfrew 1991; Barbujani et al. 1994). Such cultural factors showing the same geographical discontinuity may have acted as enhancers of the genetic separation creating a positive feed-back mechanism of differentiation and producing the demographic scenario whose genetic consequences have been detected in the present study.

Berbers from North Morocco and Algeria under Arab leadership crossed the Mediterranean sea and imposed their rule in Iberia in the 8th century AD; this lasted in the southern part of the Peninsula (i.e., Andalusia) for eight centuries, enriching the culture and bringing technological innovations. The Arab cultural influence in the northern part of the Peninsula, represented in the present study by Catalans and Basques, was far less important than in its southern part. Nonetheless, there is little differentiation between north and south of the Iberian Peninsula in the frequencies of the Alu insertions analyzed, and therefore, the different Arab cultural influence does not correlate with the genetic variation observed: Andalusians do not show shorter genetic distances to North Africans than do Catalans. The same pattern is observed in mtDNA sequences (Rando et al. 1998), but autosomal STRs (Bosch et al. 1999) show allele frequencies that may be interpreted as the result of a gene flow from NW Africa into South Iberia (Andalusians).

Basques have been shown to be a genetic outlier for autosomal markers within the European landscape (Bertranpetit and Cavalli-Sforza 1991; Calafell and Bertranpetit 1994); this has been interpreted as being the result of a European origin, an ancient divergence by drift, and later isolation. However, other authors (Martínez-Laso et al. 1995; Arnaiz-Villena et al. 1995, 1997) have proposed a common origin for Iberians (including Basques) and Berbers based on similar frequencies of some HLA haplotypes. Nonetheless, when new HLA data are added and additional numerical analyses performed, any special relationship between North African populations and Iberians compared with the rest of Europeans is not supported (Comas et al. 1998). In the present study, a weak genetic relationship has been found between Iberians and North African populations when observed in a wider framework. A common origin for both populations is not supported by the data.

Although Basques have been described as a genetic isolate within the homogeneous European genetic landscape, they fall into the European diversity throughout all the analyses performed and reported here. The frequency of Alu insertions and mtDNA (Bertranpetit et al. 1995) or Y-chromosome markers (Hurles et al. 1999; Bosch et al. 1999) do not display the clear differentiation shown by other markers (such as classical genetic markers). However, the singularity of the Basque population is shown by its extreme position in the third PC in an analysis with six European populations, and in the first and fourth PCs in an analysis with the Iberian and North African populations (data not shown). Given that the genetic differences between Basques and other Europeans probably originated by isolation and subsequent random drift (Bertranpetit and Cavalli-Sforza 1991; Calafell and Bertranpetit 1994), and that those differences are relatively small (although not so small when considered against the homogeneous European populations), it is not surprising that subsets of nuclear loci do not allow the detection of these differences. However, the 11 Alu insertion polymorphisms have enough power to resolve the stronger differences among European and North African populations.

The NW African samples analyzed present closer genetic distances to European than to sub-Saharan African populations. The present data support the hypothesis of an independent, although parallel, settlement of North Africa and South Europe (Bosch et al. 1997; Simoni et al. 1999). Both Mediterranean shores are genetically similar in comparison with sub-Saharan Africa. The Gibraltar Straits was not crossed for main settlement either from Africa into Europe or from Europe to Africa. It is known that Iberia acted as a cul de sac for Neandertals in the Middle Palaeolithic and that modern humans expanded from the Near East. It is likely that the independent establishment of anatomically modern humans on both sides of the Gibraltar Straits was the initial cause for the genetic differences. Subsequent demographic events, including the Neolithic, Mediterranean contacts (late second millennium BC to Roman empire), and the Islamic expansions appear to have had small genetic impact on north-south exchanges. It has been suggested that the Arabization of the Maghreb (NW Africa) was the result of a cultural replacement with limited demographic impact (Hitti 1990). In the present study, Berber populations are represented by samples from the Rif mountains (Northern Moroccans) and the Atlas mountain range (Southeastern Moroccans), whereas Arab populations are represented by Western Moroccans, Algerians, and Tunisians. The shortest genetic distances among all the samples tested are those between Northern and Western Moroccans, and, on the con-

trary, the genetic distances between the two Berber groups (Northern and Southeastern Moroccans) are larger. The cultural and linguistic differences shown between Arabs and Berbers are not correlated with the present data on Alu insertion polymorphisms and agree with other genetic data (Bosch et al. 1997, 1999) supporting the view that Arabization in the Maghreb was the result of a cultural replacement with little demographic impact.

In addition to the Mediterranean Sea, the other geographical barrier constraining Northern Africa, the Sahara Desert, seems to have been more permeable to human movements in its northwestern part. The finding that, in the past, the Sahara Desert was much wetter than it is nowadays might explain why the Sahara appears to have been a less significant genetic barrier than the Mediterranean. Even in northeast Africa, where the Nile could have functioned as a highway between north and south, Krings et al. (1999) have found a steep gradient of genetic differentiation in mtDNA sequences along the Nile. The shorter genetic distances observed between either Saharawi and Southeastern Moroccans and sub-Saharan populations can be explained by some admixture through the Sahara Desert. These results agree with previous genetic data (Lefranc et al. 1979; Rando et al. 1998; Bosch et al. 1999, 2000), suggesting that the admixture across the Sahara could have been achieved by continuous gene flow between both groups of populations, thus creating a cline of sub-Saharan admixture in NW Africa, which is stronger in the southern part than in the north. Nonetheless, even if the genetic influx is patent, there is a very sharp genetic differentiation between African populations north and south of the Sahara Desert. In summary, our study with Alu insertion polymorphisms shows that the Gibraltar Straits, which are only 15 km wide, have acted as a much stronger barrier to gene flow than the 2000-km-wide Sahara Desert.

Acknowledgement We thank Marjo Leppälä and Kirsti Höök for technical assistance. We also thank Lourdes Fañanás and Blanca Gutiérrez for providing the Andalusian samples, Chiheb Ben Rayana for providing some of the Tunisian samples, Noureddine Bouzerna for the collection of the Algerian samples, Marc Valls for helping us to collect the Moroccan samples, and Jordi Clarimón and Elena Bosch for their assistance in DNA extraction and for fruitful discussion. This research was supported by Dirección General de Investigación Científica y Técnica in Spain (PB95–0267-CO2–01 and PB98–1064), Direcció General de Recerca, Generalitat de Catalunya (1998SGR00009), and award number 1999-IJ-CX-K009 from the Office of Justice Programs, National Institute of Justice, Department of Justice (MAB). The points of view in this document are those of the authors and do not necessarily represent the official position of the US Department of Justice. D.C. was financially supported by the Finnish Ministry of Education and by CIMO (Centre for International Mobility) scholarships for foreign post-graduates and young researchers.

References

Arcot SS, Fontius JJ, Deininger PL, Batzer MA (1995a) Identification and analysis of a "young" polymorphic Alu element. Biochim Biophys Acta 1263:99–102

Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995b) Alu repeats: a source for the genesis of primate microsatellites. Genomics 29:136–44

Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV, Batzer MA (1996) Alu fossil relics – distribution and insertion polymorphism. Genome Res 6:1084–1092

Arnaiz-Villena A, Benmamar D, Alvarez M, Díaz-Campos N, Varela P, Gómez-Casado E, Martínez-Laso J (1995) HLA allele and haplotype frequencies in Algerians. Relatedness to Spaniards and Basques. Hum Immunol 43:259–268

Arnaiz-Villena A, Martínez-Laso J, Gómez-Casado E, Díaz-Campos N, Santos P, Martinho A, Breda-Coimbra H (1997) Relatedness among Basques, Portuguese, Spaniards and Algerians studied by HLA allelic frequencies and haplotypes. Immunogenetics 47:37–43

Barbujani G, Pilastro A, De Domenico S, Renfrew C (1994) Genetic variation in North Africa and Eurasia: Neolithic demic diffusion vs Paleolithic colonisation. Am J Phys Anthrop 95: 137–154

Batzer MA, Deininger PL (1991) A human-specific subfamily of Alu sequences. Genomics 9:481–487

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994) African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91: 12288–12292

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M (1996) Genetic variation of recent Alu insertions in human populations. J Mol Evol 42:22–29

Bertranpetit J, Cavalli-Sforza LL (1991) A genetic reconstruction of the history of the population of the Iberian Peninsula. Ann Hum Genet 55:51–67

Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995) Human mitochondrial DNA variation and the origin of Basques. Ann Hum Genet 59:63–81

Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, Bertranpetit J (1997) Population history of North Africa: evidence from classical genetic markers. Hum Biol 69:295–311

Bosch E, Calafell F, Santos FR, Pérez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) STR variation is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623–1638

Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, Mateu E, Martínez-Arias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Hariti G, Cambon-Thomsen A, Bertranpetit J (2000) Genetic structure of northwestern Africa revealed by STR analysis. Eur J Hum Genet 8:360–366

Calafell F, Bertranpetit J (1994) Principal component analysis of gene frequencies and the origin of Basques. Am J Phys Anthropol 93:201–215

Comas D, Mateu E, Calafell F, Pérez-Lezaun A, Bosch E, Martínez-Arias R, Bertranpetit J (1998) HLA class I and class II DNA typing and the origin of Basques. Tissue Antigens 51: 30–40

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes:application to human mitochondrial DNA restriction data. Genetics 131:479–491

Felsenstein J (1989) PHYLIP – phylogeny inference package. Cladistics 5:164–166

Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Hitti P (1990) The Arabs: a short history. Gateway, Washington DC

Hurles ME, Veita R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, López de Munain A, Röhl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 65:1437–1448

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999) MtDNA analysis of Nile river valley populations: a genetic corridor or a barrier to migration? Am J Hum Genet 64:1166–1176

Lefranc G, Lange G de, Rivat L, Langaney A, Lefranc MP, Ellouze F, Sfar G, Sfar M, Loghem E van (1979) Gm, Am and Km immunoglobulin allotypes of two populations in Tunisia. Hum Genet 50:199–211

Martínez-Laso J, De Juan D, Martínez-Quiles N, Gómez-Casado E, Cuadrado E, Arnaiz-Villena A (1995) The contribution of the HLA-A, -B, -C and –DR, -DQ DNA typing to the study of the origins of Spaniards and Basques. Tissue Antigens 45: 237–245

Newman JL (1995) The peopling of Africa. A geographic interpretation. Yale University Press, New Haven London

Novick GE, Novick CC, Yunis J, Yunis E, Antúnez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ (1998) Polymorphic Alu insertions and the Asian origin of Native American populations. Hum Biol 70: 23–39

Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt HJ (1998) Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. Ann Hum Genet 62:531–550

Renfrew C (1991) Before Babel: speculations on the origins of linguistic diversity. Cambridge Archaeol J 1:3–23

Reynolds J, Weir BS, Cockerman CC (1983) Estimation of the coancestry coefficient: basis for a short term genetic distance. Genetics 105:767–779

Said R, Faure H (1990) Chronological framework: African pluvial and glacial epochs. In: Ki-Zerbo J (ed) General history of Africa. UNESCO, Paris, pp 146–166

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schneider S, Kueffer JM, Roessli D, Excoffier L (1996) Arlequin (ver. 1.0): a software enviroment for the analysis of population genetics data. Genetics and Biometry Lab, University of Geneva, Geneva, Switzerland

Simoni L, Gueresi P, Pettener D, Barbujani G (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. Hum Biol 71:399–415

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071