



ARTICLE

# Genetic structure of north-west Africa revealed by STR analysis

Elena Bosch<sup>1</sup>, Francesc Calafell<sup>1</sup>, Anna Pérez-Lezaun<sup>1</sup>, Jordi Clarimón<sup>1</sup>, David Comas<sup>1</sup>, Eva Mateu<sup>1</sup>, Rosa Martínez-Arias<sup>1</sup>, Bernal Morera<sup>1</sup>, Zahra Brakez<sup>2</sup>, Omar Akhayat<sup>2</sup>, Abdelaziz Sefiani<sup>3</sup>, Ghania Hariti<sup>4</sup>, Anne Cambon-Thomsen<sup>5</sup> and Jaume Bertranpetit<sup>1</sup>

<sup>1</sup>Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Spain; <sup>2</sup>Laboratoire de Biologie Cellulaire et Moléculaire, Faculté des Sciences, Université Ibnou Zohr, Agadir; <sup>3</sup>Institut National d'Hygiène, Rabat, Morocco; <sup>4</sup>Hôpital Mustapha, CHU Alger Centre, Algeria; <sup>5</sup>INSERM U 518, Faculté de Médecine, Toulouse, France

We have analysed a large set of autosomal short tandem repeat (STR) loci in several Arabic and Berber-speaking groups from north-west Africa (ie Moroccan Arabs, northern-central and southern Moroccan Berbers, Saharawis, and Mozabites). Two levels of analysis have been devised using two sets of 12 STR loci, (D3S1358, vWA, FGA, THO1, TPOX, CSF1PO, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820) and 21 (the former set plus D9S926, D11S2010, D13S767, D14S306, D18S848, D2S1328, D4S243, F13A1, and FES/FPS). For each set, data for a number of external reference populations were gathered from the literature. Several methods of analysis based on genetic distances (neighbour-joining trees, principal coordinate analysis, boundary detection), as well as AMOVA, showed that genetic differentiation among NW African populations was very low and devoid of any spatial pattern. When the NW African populations were grouped according to cultural or linguistic differences, the partition was not associated with genetic differentiation. Thus, it is likely that Arabisation was mainly a cultural process. A clear genetic difference was found between NW African populations and Iberians, which underscores the Gibraltar Straits as a strong barrier to genetic exchange; nonetheless, some degree of gene flow into Southern Iberia may have existed. NW Africans were genetically closer to Iberians and to other Europeans than to African Americans. *European Journal of Human Genetics* (2000) 8, 360–366.

**Keywords:** STRs; microsatellites; population genetics; North Africa

## Introduction

North-west Africa is host to a number of populations that are different in their language, culture and ethnic identity. We have studied several Arabic and Berber-speaking populations from NW Africa by means of a large set of autosomal short tandem repeat (STR), also known as microsatellite loci in an attempt to genetically characterise that region. Berber is a branch of the Afro-Asiatic language family; this branch embraces more than 30 distinct languages with a distribution

from Egypt to Senegal which were the only languages spoken in the area until the Arab invasion. Arabic is a language classified within the Semitic branch of the Afro-Asiatic family and includes several regional dialects.<sup>1</sup> Arabic was introduced into NW Africa from the 7th century by the Moslem invasions and is spoken nowadays by the Moroccan Arab population. Saharawis, who live in the Western Sahara, speak Hassani, a different Arabic dialect. Tachelhit, spoken mainly in the Souss valley, is a Berber language confined to southern Morocco and distinct from other Berber languages spoken in the northern (Tarifit) and central (Tamazigh) areas of Morocco. The Mozabites are a very well defined Berber population in Algeria: they speak Mzab (a distinct Berber language) and originated from the Ibadite religious community who settled in this region in the 11th century.

Correspondence: Jaume Bertranpetit, Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003- Barcelona, Spain. Tel: +34 93 542 28 40; Fax: +34 93 542 28 02; E-mail: jaume.bertranpetit@cexs.upf.es  
Received 15 July 1999; revised 12 December 1999; accepted 6 January 2000

We are interested in elucidating the possible structure of STR variation in these NW African populations, but also in knowing about their relations with neighbouring populations to the north – the Iberians, and to the south, the Sub-Saharan Africans. Each unlinked autosomal STR behaves as an independent locus; the study of a number of unlinked autosomal STRs gives insight into a number of independent effects of the evolutionary process. Analysis of these elements may reveal overall trends that may have been obscured by stochastic factors at any single locus.

The aims of this study are to approach:

- (1) The heterogeneity of speakers of different Berber dialects and languages. These are the North African human groups more likely to have preserved ancient characteristics as they preserve the oldest known language and social structure. The extent of genetic differentiation and of the loss of diversity within these groups can be used to ascertain the historical patterns of isolation or gene flow between these populations;
- (2) The genetic differentiation between Arabs and non-Arabs. The apportionment of genetic diversity by ethnicity in NW Africa (Arab vs Berbers and Saharawis) can provide information about the demographic impact of the Arab expansions in the late seventh and the 11th centuries AD in the autochthonous populations of the region. Although there is clear cultural and linguistic differentiation, it is not clear at all to what extent a main demographic migration caused the cultural change or whether it was acculturation without genetic replacement;
- (3) The extent of population exchange associated with migrations between NW Africa and Iberia. No significant historical movements are known from the Iberian Peninsula to NW Africa, except for the expulsions of the Jews in the late 15th century and of the Moors in the 17th century. The Islamic invasion of the Iberian Peninsula in 711 AD (with subsequent invasions in the 11th century) appears to be the main migration event, although its actual demographic impact may have not been large.<sup>2,3</sup> In our study, we have compiled Iberian data from the literature that covers areas with none or brief Arab occupation (Basques, Catalans and northern Portuguese) as well as samples from areas with a much longer Arab occupation (Andalusians).

## Materials and methods

### Population samples analysed

The NW African samples analysed in the present study include a Moroccan Arab population ( $2n = 94$ – $160$ ), a pooled group of Berber speakers from northern and central Morocco ( $2n = 50$ – $126$ ), a Berber population from southern Morocco ( $2n = 84$ – $96$ ), a Saharawi population sample ( $2n = 104$ – $118$ )

from the Western Sahara plus an Algerian Berber group, the Mozabites, collected in the town of Ghardaia ( $2n = 88$ ). Differences in number of chromosomes analysed within a population were due to DNA availability and to the different methods used for typing each STR. Geographic location is shown in Figure 3. Allele frequencies for 13 loci in a partial subsample of Arabs and northern Berbers are given by Pérez-Lezaun *et al.*<sup>4</sup> DNA was extracted from fresh whole blood using standard phenol–chloroform methods. Appropriate informed consent was obtained from all participants in this study and in most cases information about geographic origin and native language of their four grandparents was recorded.

### STR typing

Most samples were PCR amplified using commercial kits (PE Applied Biosystems, Foster City, CA, USA) AmpF/STR Profiler Plus (D3S1358, VWF/VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820) and AmpF/STR Cofiler (D3S1358, D16S539, Amelogenin, TH/THO1, TPOX, CSF1PO and D7S820) or AmpF/STR Green I (Amelogenin, TH/THO1, TPOX and CSF1PO) according to manufacturers' recommendations. Additional samples were genotyped for some STR loci (VWF/VWA, D5S818, D7S820, TH/THO1, TPOX) overlapping the commercial kits using fluorescently labelled primers. Nine other STR loci (D11S2010, D13S767, D14S306, D18S848, D2S1328, D4S243, F13A1, FES/FPS, D9S926) were typed in all the NW African populations except for the Mozabites, as no sufficient DNA was available, also using fluorescently labelled primers. Almost all loci map in different chromosomes. For primer sequences see Pérez-Lezaun *et al.*<sup>4</sup> Single locus amplifications were carried out in a final reaction volume of  $10\ \mu\text{l}$  according to the following touchdown PCR cycling conditions:  $94^\circ\text{C}$  for 1 min, 14 cycles at  $94^\circ\text{C}$  for 20 s, with annealing temperatures decreasing from  $63^\circ\text{C}$  by 0.5 degrees C in each cycle for 1 min, and  $72^\circ\text{C}$  for 1 min, 20 cycles at  $94^\circ\text{C}$  for 20 s,  $56^\circ\text{C}$  for 45 s, and  $72^\circ\text{C}$  for 1 min, plus a final extension cycle at  $72^\circ\text{C}$  for 5 min. Amplification products were run in an ABI 377<sup>TM</sup> sequencer, and ABI GS500 ROX and ABI GS500 TAMRA was used as internal lane standards. The GeneScan 672<sup>TM</sup> and Genotyper 2.1 x 3<sup>TM</sup> software packages were used to collect the data, analyse fragment sizes and to designate alleles by comparison with locus-specific ladders.

### Reference populations

Original references for Catalans ( $2n = 100$ – $176$ ) and Basques ( $2n = 98$ – $200$ ) can partly be found in an overall European sample that comprised two other European samples.<sup>4,5</sup>

There are original references for Portuguese ( $2n = 72$ ) and Andalusians ( $2n = 68$ – $72$ )<sup>4</sup> and Italians ( $2n = 446$ ).<sup>6</sup> Population data for African Americans ( $2n = 390$ ) and European Americans ( $2n = 400$ , also called Caucasians in the original reference) is available in AmpF/STR Profiler Plus and AmpF/

STR Cofiler (PE Applied Biosystems, Foster City, CA, USA) User's Manuals.

### Statistical analysis

In a compromise between the number of populations and the number of STR loci that have been typed in them, we devised two levels of analysis using two different data sets, which we call basic and extended. In the basic data set, we use 12 STR loci (D3S1358, vWA, FGA, THO1, TPOX, CSF1PO, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820) typed in all the NW African populations analysed in this study as well as in all reference populations cited above. In the extended data set, we use 21 STR loci (that consist of the 12 included in the basic data set plus D9S926, D11S2010, D13S767, D14S306, D18S848, D2S1328, D4S243, F13A1, and FES/FPS) typed in Moroccan Arabs, northern central and southern Moroccan Berbers, and Saharawis, and that are also available in the literature for two populations from the Iberian Peninsula (Basques and Catalans).

We computed  $F_{st}$  distances<sup>7</sup> between each pair of populations from the basic and extended data sets and represented them as neighbour-joining trees.<sup>8</sup> Neighbour-joining tree robustness was assessed by bootstrap analysis;<sup>9</sup> every occurrence of a particular cluster was recorded and given as a percentage of the 10000 bootstrap trees drawn from the previously bootstrapped matrix distances. Principal coordinate analysis was performed on the  $F_{st}$  distance matrix by using NTSYS-pc version 1.70 (Applied Biostatistics, Inc., Setauket, NY, USA).

A Delaunay network<sup>10</sup> was used to define pairs of contiguous samples. In this way, the nine estimated localities for the NW African and Iberian populations used in the basic data set were connected by 17 edges, so the edges define around each locality the polygon that contains the set of points that is closer to that locality than to any other. We associated with each edge the genetic distance between the pair of populations it links. To identify the zones of sharpest genetic change, or genetic boundaries, we initially traced a perpendicular line across the edge showing the highest genetic distance, which was the origin of the first boundary. The boundary was then extended across the adjacent edges showing the highest genetic distances until it reached the limits of the network. This procedure can be iterated to define a second, a third or any higher order boundary.

Analysis of molecular variance or AMOVA<sup>11</sup> was performed using Arlequin v.1.1<sup>12</sup> independently for each locus; the resulting genetic variance apportionment fractions were averaged over loci and their associated  $P$  values were combined by the Fisher technique.<sup>13</sup>

## Results

We studied north-west African populations by typing 21 autosomal STR loci in Moroccan Arabs, northern central and southern Moroccan Berbers, Saharawis, and a subset of

12 loci in Mozabites. Allele frequencies for each locus and population were part of the review process of this paper and are available on request from the authors (francesc.calafell@cexs.upf.es). Hardy-Weinberg equilibrium was tested for all possible locus–population combinations by two methods:  $\chi^2$  comparison of observed vs expected homozygotes, and by an exact test as implemented in the Arlequin package. Three and seven, respectively, out of 96  $\chi^2$  and exact tests were statistically significant ( $P < 0.05$ ). No consistent deviations were observed at several loci in a single population, nor at a single locus in most populations; no significant deviations remained after Bonferroni correction for multiple testing. Therefore, equilibrium may be assumed for all loci in all populations.

### STR diversity within populations

Within each population of the basic and extended sets, STR variability has been explored by means of both mean expected heterozygosity and mean allele length variance per locus (Table 1). In the basic set, we found levels of heterozygosity that were very similar across all the NW African populations (Kruskal-Wallis test,  $P = 0.828$ ). Only the Mozabite population seemed to present a slight reduction in heterozygosity with a mean value of  $0.763 \pm 0.016$ , although the differences from the other populations were not statistically significant (Mann Whitney's U,  $P = 0.237$ ). When compared with the external populations, heterozygosities in overall NW Africa ( $0.782 \pm 0.016$ ) were very similar to those in Iberians ( $0.780 \pm 0.017$ ) and other Europeans ( $0.784 \pm 0.019$ ) (NW African vs the remaining populations, Mann-Whitney's U,  $P = 0.672$ ). African-Americans, with a heterozygosity of  $0.793 \pm 0.015$ , were only slightly more diverse. Estimates in the extended set yielded a very similar pattern.

Variance in allele length per locus displayed the same pattern as heterozygosity: slight reduction in the Mozabites ( $4.64 \pm 1.37$ ) with no statistically significant differences (Mann Whitney's U,  $P = 0.605$ ) across the other NW Africans, and similar levels of genetic variation (Kruskal-Wallis,  $P = 0.929$ ) in overall NW Africans ( $5.31 \pm 1.62$ ), Europeans ( $5.08 \pm 1.41$ ), and African Americans ( $5.67 \pm 1.90$ ).

### STR diversity among populations

Genetic differentiation among the populations of both the extended and basic sets was analysed by computing  $F_{st}$  genetic distances and representing them by means of neighbour-joining trees (Figures 1a and 1b). We also computed other genetic distances (data not shown): almost identical tree topologies were obtained with  $D_{sw}$ <sup>14</sup> and with  $(\delta\mu)^2$ ,<sup>15</sup> whereas  $R_{ST}$ <sup>16</sup> yielded a poorer tree, in accordance with previous studies comparing these four distance measures.<sup>17</sup> In both  $F_{st}$  trees (Figures 1a and 1b), the NW African populations cluster together, although with short and not very statistically robust branches among them. In the basic analysis, the Mozabites stand out from the rest of the NW

Africans. As stated above, Mozabites have less internal diversity than other NW African populations. Both facts seem to indicate that they may have differentiated by drift. It should be noted that the most robust branch (77%) in the tree is that separating NW Africans from Europeans. Among the Iberians, Andalusians and Portuguese are closest to NW Africa. African-Americans appear linked to NW Africa through a long branch. In the extended data analysis, we find again little structure among NW Africans although with higher percentages of bootstrap support and a very robust separation from Iberians with a bootstrap value of 97%.

We also have represented the distance matrix by principal coordinate analysis (Figure 2). Together, the first two principal coordinates account for 75% of the variance in the distance matrix. The genetic picture obtained in the two-dimensional representation (Figure 2) is similar to that displayed in the neighbour-joining topologies: all NW Africans, except for Mozabites, are close to each other and separated from Europeans and African-Americans. The only discordance between the neighbour-joining topologies and the principal coordinate analysis is found in the Andalusians, who in this new representation hardly depart from the European genetic pool.

The two strongest genetic boundaries in the geographical area comprised by NW Africa and the Iberian Peninsula (see Figure 3) were found to encircle single populations, the Mozabites and the Basques. The third genetic boundary separated NW Africa from the Iberian Peninsula and finally, the fourth genetic boundary encircles the Saharawis.

#### Apportionment of genetic variance

We also explored how genetic variance is apportioned among different geographic, cultural or linguistic population groups by means of the analysis of molecular variance (AMOVA). Within NW Africa, only 0.36% ( $P = 0.0009$ ) of the genetic variance in the basic data set is found between populations, while the rest is found within them. In the extended data set (which does not contain the Mozabites), the fraction of the

genetic variance found among NW African populations dropped to 0.05% ( $P = 0.147$ ). When NW African populations are divided into Moroccan Arabs and non-Arabs (northern central and southern Moroccan Berbers, Saharawis and Mozabites), the genetic variance attributable to this partition is not significantly different from zero (0.13%,  $P = 0.590$  for the basic data set and 0.19%,  $P = 0.633$  for the extended data set). When we compared the NW Africans with the Iberian populations, the fraction of genetic variance attributable to the difference between these two groups was 0.54% ( $P = 0.0008$ ) for the basic data set and 0.75% ( $P = 0.001$ ) for the extended data set. These values are comparable to those found between NW Africans and non-Iberian Europeans (Italians and European Americans, 0.93%,  $P < 0.0001$ ) for the basic data set.

#### Discussion

We have typed a large set of autosomal STR loci in several Berber and Arabic speaking population groups in order to obtain a genetic characterisation of north-west Africa. Different analyses have been performed to test several hypotheses about the population history in this geographical region.

The most distinctive feature of our results is the substantial genetic similarity found among most of the NW African populations studied. Although they represent different cultural and/or ethnical population groups, there is no clear pattern of genetic differentiation between cultural or ethnic groups. Given their cultural differences, the three Berber-speaking groups studied here cannot be considered as a single homogenous group. However, we find low levels of genetic differentiation among them. The Mozabites, which are geographically distant from all other populations considered, may be an exception. In this case, they seem to be the most distinctive and genetically isolated population within NW Africa. No significant genetic differences were found between Arabs and non-Arabs (ie Berbers and Saharawis). Another significant finding in this study concerns the comparison of

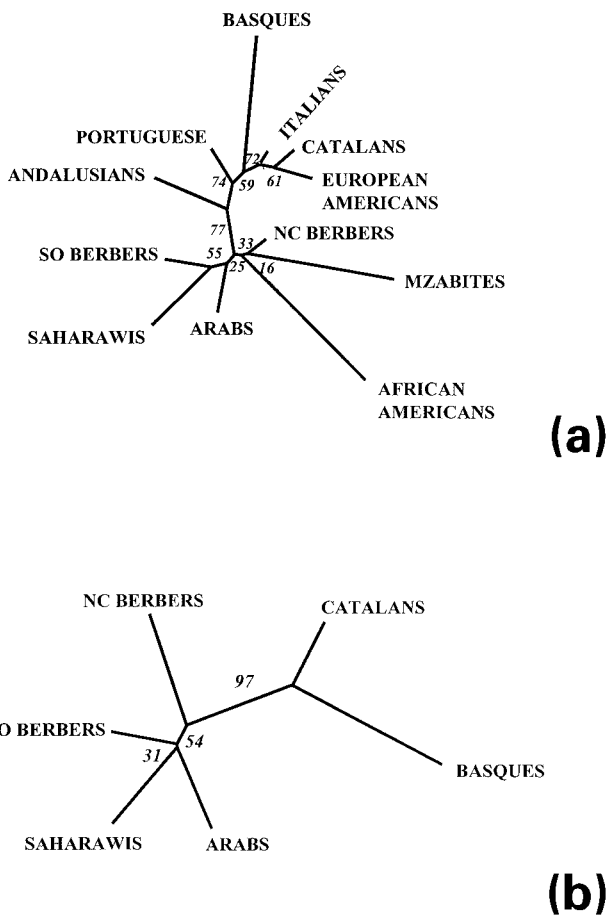
**Table 1** Mean expected heterozygosity and allele length variance per locus ( $\pm$  standard error) for the populations in the basic and extended data sets

	Mean expected heterozygosity		Mean allele length variance	
	basic	extended	basic	extended
ARA	0.790 $\pm$ 0.016	0.758 $\pm$ 0.015	5.74 $\pm$ 1.76	4.20 $\pm$ 1.10
SAH	0.782 $\pm$ 0.019	0.753 $\pm$ 0.015	5.30 $\pm$ 1.59	3.87 $\pm$ 0.99
SOB	0.788 $\pm$ 0.017	0.761 $\pm$ 0.013	5.44 $\pm$ 1.63	3.97 $\pm$ 1.01
NCB	0.786 $\pm$ 0.017	0.745 $\pm$ 0.019	5.42 $\pm$ 1.73	3.90 $\pm$ 1.08
MZA	0.763 $\pm$ 0.016	—	4.64 $\pm$ 1.37	—
CAT	0.782 $\pm$ 0.017	0.746 $\pm$ 0.016	5.01 $\pm$ 1.43	3.58 $\pm$ 0.90
BAS	0.775 $\pm$ 0.018	0.736 $\pm$ 0.016	5.72 $\pm$ 1.62	3.96 $\pm$ 1.02
POR	0.782 $\pm$ 0.019	—	5.05 $\pm$ 1.45	—
AND	0.779 $\pm$ 0.015	—	4.54 $\pm$ 1.21	—
ITA	0.784 $\pm$ 0.019	—	5.07 $\pm$ 1.37	—
EAM	0.779 $\pm$ 0.022	—	5.04 $\pm$ 1.62	—
AAM	0.793 $\pm$ 0.015	—	5.64 $\pm$ 1.90	—

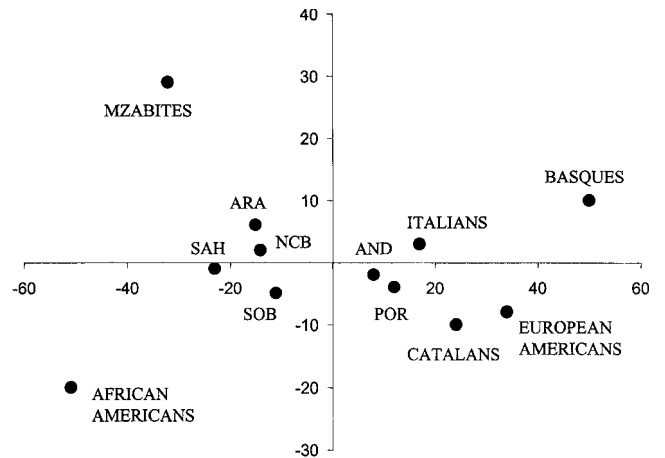
ARA: Moroccan Arabs; SAH: Saharawis; SOB: southern Moroccan Berbers; NCB: northern central Moroccan Berbers; MZA: Mozabites; CAT: Catalans; BAS: Basques; POR: Portuguese; AND: Andalusians; AAM: African-Americans; EAM: European-Americans; ITA: Italians.

the NW African autosomal gene pool with that of different external reference populations. The two neighbour-joining trees, analysis of principal coordinates, and analysis of the main genetic boundaries point to the grouping of all the NW African populations in a cluster quite separated and differentiated from those of the Iberian Peninsula and other populations within the European genetic variability. Nonetheless, the difference between the two continents is lower than the differentiation of some specific populations: Basques in Iberia and Mozabites in NW Africa.

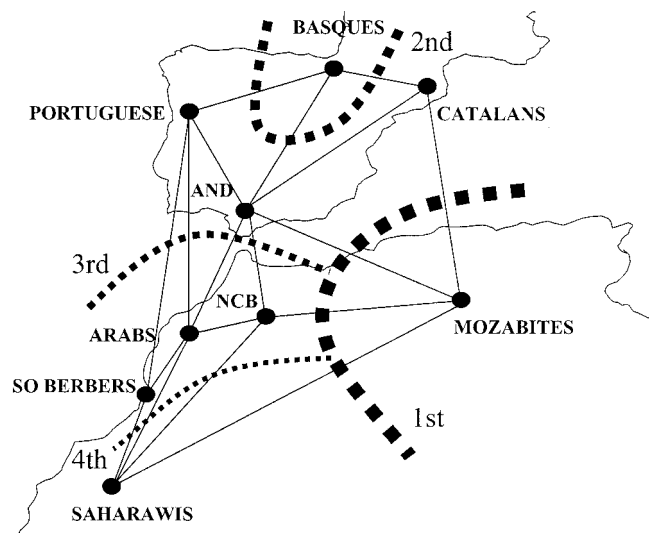
As discussed above, the Mozabites present slightly reduced heterozygosities, although no statistically significant differences were found. Heterozygosity in a population can be related to its effective population size and degree of genetic isolation. This is in agreement with the history and way of life of the Mozabites who are a religious community that



**Figure 1** Neighbour-joining trees based on  $F_{st}$  genetic distance for the populations and loci included in the basic (a) and extended (b) data set. The figures along the tree branches represent the percentage of times that a certain branch is found in 10000 bootstrapped trees. Arabs: Moroccan Arabs; NC Berbers: northern central Moroccan Berbers; SO Berbers: southern Moroccan Berbers.



**Figure 2** Representation of the two first principal coordinate scores derived from the analysis of the extended  $F_{st}$  genetic distance matrix. All score values on the coordinates have been multiplied by a factor of  $10^5$ . ARA: Moroccan Arabs; NCB: northern central Moroccan Berbers; SOB: southern Moroccan Berbers; SAH: Saharawis; AND: Andalusians; POR: Portuguese.



**Figure 3** Genetic boundaries found by superimposing an  $F_{st}$  distance matrix on a Delaunay triangulation. AND: Andalusians; NCB: northern central Moroccan Berbers; SO Berbers: southern Moroccan Berbers.

obeys a strict rule of the Ibadite tradition and protects its identity by a socially prescribed strong endogamy.<sup>18</sup>

The  $F_{st}$  values obtained among the NW African populations are low and without a clear structure, as shown by the star-like shape of the neighbour-joining tree, with short internal branches among the NW African populations. This observation agrees with the low fraction of genetic variation found among NW African populations, 0.36%, which is about one order of magnitude less than that obtained in a previous study with classical polymorphisms covering a larger area.<sup>19</sup>

A possible cause of the high observed homogeneity may lie in the specific properties of loci analysed. We have used in this study several STR loci included in the commercial kits AmpF/STR Profiler Plus, AmpF/STR Cofiler and AmpF/STR GreenI, which contain the marker set designed for forensic databanking in the USA. Among other desirable properties, it may be the case that some loci included in the commercial kits showed similar allele frequency distributions among 'European' subpopulations. Thus, it is possible that some of our results (such as the absence of a clear genetic structure among the NW African populations) could be biased by the markers selected as a core set in US forensic genetics which constitute the commercial kits used. In order to test this possible bias, we computed the average  $F_{st}$  of the NW African populations considered in the extended data set, separately for the STR loci used in the commercial kits and for those STR not included in the kits (see Materials and methods). The means obtained were, respectively,  $0.0112 \pm 0.0010$  and  $0.0104 \pm 0.0013$ , which were not statistically significantly different (Mann-Whitney's U,  $P = 0.345$ ). We can conclude that the lack of differentiation among the Berber and Arabic-speaking populations in NW Africa cannot be due to possible biases in the commercial STR set used and we may reject a spurious cause of our observation of high genetic homogeneity in NW Africa. Given that the linguistic and cultural differences among Arabic and Berber speakers is not reflected by their genes, it is a plausible argument that the Arabisation of NW Africa was only a cultural phenomenon with subsequent little genetic impact.

We have found quite clear genetic differentiation between Iberians and NW Africans in several of the analysis presented. Both groups of populations seem to belong to two different genetic pools. However, it should be noted that they are closer to each other than to non-European groups, as depicted by the large genetic distance to African-Americans. This observation is compatible with the hypothesis suggested elsewhere.<sup>19</sup> NW Africa was peopled in the Upper Paleolithic era with anatomically modern humans roughly at the same time that the same human groups colonised Europe. Subsequently, the ancestors of the Berbers differentiated *in situ* before the Neolithic wave of advance,<sup>20</sup> and this demic diffusion, which seemed to have a large impact on the genetic make up of Europeans,<sup>21</sup> may have had little impact in NW Africa.

This general genetic differentiation between NW Africa and the Iberian Peninsula does not apply equally to all populations. While some Iberians have genetic distances to NW Africans that are as large as those between other Europeans and NW Africans, others, particularly the Andalusians, seem to be genetically closer and to have received some gene flow from NW Africa. The possible genetic contribution of NW Africa to the Andalusian population may be estimated by using the triangle method.<sup>22</sup> This estimate seems to point to an important contribution of NW Africa to the Andalusian gene pool ( $m = 0.317$ , 95% confidence inter-

val from  $-0.006$  to  $0.802$ ), although it is not significantly different from zero. This result should be taken with caution given the statistical uncertainty and the lack of knowledge about the genetic make-up of the ancestral populations. Genetic systems for which the gene genealogy can be traced are better suited for this purpose: a group of mtDNA sequences (called U6)<sup>23</sup> seems to be found almost only in NW Africa, where it reaches frequencies up to 25%. U6 sequences were found in three out of 54 Portuguese<sup>24</sup> and in two out of 92 Galicians,<sup>25</sup> but were absent in a sample of 15 Andalusians<sup>24</sup> and in 162 other Iberians.<sup>24,26,27</sup> Thus, when we combine autosomal, Y chromosome (Bosch *et al*, in preparation) and mtDNA data,<sup>23</sup> we may conclude that NW African gene flow into southern Iberia seems to be small.

As expected, Basques appeared as the most differentiated population within the Iberian Peninsula. Some authors<sup>28,29</sup> studied HLA variation, and suggested, based on similar frequencies for some haplotypes, that Basques and Berbers had a common origin. This conclusion was challenged by new data and a proper numerical analysis.<sup>30</sup> The present results, based on up to 21 autosomal STRs, argue against this hypothetical genetic similarity; caution must be taken when making population inferences based on a single locus.

North African population genetics seem to present interesting challenges that make it a case study for understanding how a complex population history has shaped the present gene pool.

#### Acknowledgements

We thank all the blood donors who made this study possible. This research was supported by Dirección General de Investigación Científica y Técnica in Spain (projects PB95-0267-CO2-01 and PB98-1064) and by Comissionat per a Universitats i Recerca, Generalitat de Catalunya (1998SGR00009). The latter agency also supported EB (FI/96-1153) and Spanish Ministry of Education and Science supported RM-A (AP96). The Mozabite samples from Ghardaia were obtained in the framework of an INSERM network: 490NS1 under the coordination of Pr MS Isaad (Algiers) and with the help of the medical team of Ghardaia. The help of JM Dugoujon and A Sevin, CNR Toulouse, is gratefully acknowledged. We especially thank all persons involved in reaching the Saharawi donors as well as Elisabeth Pintado (Sevilla), Josep Lluís Fernández Roure and Alba Bosch (Mataró) for their help in contacting Moroccan donors. We appreciate criticism of our work by Michael Krawczak.

#### References

- 1 Ruhlen M: *A Guide to the World's Languages*, 2nd edn. Stanford University Press: Stanford, 1991.
- 2 Sánchez-Albornoz C: *El Islam de España y el Occidente*. Centro italiano di studi sull'alto medioevo: Spoleto, 1965.
- 3 Guichard P: Structures sociales orientales et occidentales dans l'Espagne musulmane. Mouton: Paris, 1977.
- 4 Pérez-Lezaun A, Calafell F, Clarimón J *et al*: Allele frequencies of 13 short tandem repeats in population samples of the Iberian Peninsula and Northern Africa. *Int J Legal Med* (in press).
- 5 Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Bertranpetit J: Allele frequencies for 20 microsatellite loci in a worldwide population survey. *Hum Hered* 1997; 47: 189-196.

- 6 Garofano L, Pizzamiglio M, Vecchio C *et al*: Italian population data in thirteen short tandem repeat loci: HUMTHO1, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358. *For Sci Int* 1998; **97**: 53–60.
- 7 Reynolds J, Weir BS, Cockerham CC: Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* 1983; **105**: 767–779.
- 8 Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**: 406–425.
- 9 Felsenstein J: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**: 783–791.
- 10 Brassel KE, Reif D: A procedure to generate Thiessen polygons. *Geogr Anal* 1979; **11**: 289–303.
- 11 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.
- 12 Schneider S, Kueffer J-M, Roessli D, Excoffier L: *Arlequin vers. 1.1: A software for Population Genetic Data Analysis*. Genetics and Biometry Laboratory, University of Geneva: Switzerland, 1997.
- 13 Sokal RR, Rohlf FJ: *Biometry*. Freeman: New York, 1995.
- 14 Shriver MD, Jin L, Boerwinkle E, DeKra R, Ferrell RE, Chakraborty R: A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* 1995; **12**: 914–920.
- 15 Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW: An evaluation of genetic distances for use with microsatellite loci. *Genetics* 1995; **139**: 463–471.
- 16 Slatkin M: A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 1995; **139**: 457–462.
- 17 Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J: Microsatellite variation and the differentiation of modern humans. *Hum Genet* 1997; **99**: 1–7.
- 18 Khaldoun I: *Histoire des Berbères et des Dynasties Musulmanes de l'Afrique Septentrionale*. Paul Geuthner (ed.), Librairie Orientaliste: Alger, 1956.
- 19 Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, Bertranpetit J: A population history of Northern Africa: evidence from classical genetic markers. *Hum Biol* 1997; **69**: 295–311.
- 20 Ammerman AJ, Cavalli-Sforza LL: *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press: Princeton, 1984.
- 21 Menozzi P, Piazza A, Cavalli-Sforza LL: Synthetic maps of human gene frequencies in Europeans. *Science* 1978; **201**: 786–792.
- 22 Cavalli-Sforza LL, Menozzi P, Piazza A: *History and Geography of Human Genes*. Princeton University Press: Princeton, 1994.
- 23 Rando JC, Pinto F, Gonzalez AM *et al*: Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann Hum Genet* 1998; **62**: 531–550.
- 24 Côte-Real HBSM, Macaulay V, Richards MB *et al*: Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 1996; **60**: 331–350.
- 25 Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A: mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 1998; **6**: 365–375.
- 26 Bertranpetit J, Sala J, Calafell F, Underhill P, Moral P, Comas D: Human mitochondrial DNA variation and the origin of the Basques. *Ann Hum Genet* 1995; **59**: 63–81.
- 27 Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM: Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* 1996; **60**: 321–330.
- 28 Arnáiz-Villena A, Benmamar D, Alvarez M *et al*: HLA allele and haplotype frequencies in Algerians. Relatedness to Spaniards and Basques. *Hum Immunol* 1995; **43**: 259–268.
- 29 Arnáiz-Villena A, Martínez-Laso J, Gómez-Casado E *et al*: Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes. *Immunogenetics* 1997; **47**: 37–43.
- 30 Comas D, Mateu E, Calafell F *et al*: HLA class I and class II DNA typing and the origin of Basques. *Tissue Antigens* 1998; **51**: 30–40.