doi: 10.1046/j.1529-8817.2005.00170.x

Temporal Mitochondrial DNA Variation in the Basque Country: Influence of Post-Neolithic Events

A. Alzualde[1], N. Izagirre[1], S. Alonso[1], A. Alonso[2] and C. de la Rúa[1,*]

[1]*Euskal Herriko Unibertsitatea (UPV/EHU), Zientzia eta Teknologia Fakultatea. Genetika, Antropologia Fisikoa eta Animali Fisiologia Saila. Posta Kutxa 644, 48080 Bilbo, Spain*

[2]*Instituto Nacional de Toxicología y Ciencias Forenses, Servico de Biología. Luis Cabrera, 9 28002 Madrid, Spain*

Summary

The Basque population has been considered an outlier in a large number of genetic studies, due to its hypothesized antiquity and greater genetic isolation. The present paper deals with an analysis of the mtDNA variability of the historical population of Aldaieta (VI–VII c. AD; Basque Country) which, together with genetic data existing for other prehistoric populations of the Basque Country (4,500–5,000 YBP), permits an appraisal of the hypotheses proposed for the origin of the genetic differentiation of the Basque population. Given that this is an aDNA study, application has been made both of standard precautions, to avoid contamination, and of authentication criteria (analysis of duplicates, replication in an independent laboratory, quantification of target DNA, sequencing and cloning of PCR products). The variability of the mtDNA haplogroups of the historical population of Aldaieta falls within the range of the present-day populations of Europe's Atlantic fringe, whereas the prehistoric populations of the Basque Country display clear differentiation in relation to all others. Consequently, we suggest that between 5,000–1,500 YBP approximately, there may have been gene flow amongst the western European populations that homogenised mtDNA lineages.

Keywords: ancient DNA, mitochondrial DNA, Basque population, post-Neolithic gene flow

Introduction

On the basis of archaeological data, inference has been forthcoming of several large-scale demographical events that occurred in Europe: the continent's colonisation by anatomically modern humans from the Near East (Straus L.G., 1989; Mellards, 1992); the re-expansion of the population from glacial refugia (Southern Pyrenees, Alps and Balkans) towards Northern Europe following the Last Glacial Maximum (Otte, 1990); and the introduction of agriculture from the Near East towards Western and Eastern Europe (Ammermann & Cavalli-Sforza, 1984). There appears to be no other documented event that has had an overall bearing on European demography, so that genetic patterns at a European level tend to be interpreted within the context of any one of the aforementioned events (Barbujani & Bertorelle, 2001).

In an initial paper based on the frequency of classic genetic markers amongst European and Near Eastern populations, undertaken by Menozzi *et al*. (1978), a SE-NW gradient was observed which coincided exactly with the archaeological map of Neolithic expansion, whereby the study provided support for the model of Neolithic demic diffusion proposed by Ammerman & Cavalli-Sforza (1973). There are authors who consider that the SE-NW gradient identified by Menozzi *et al*. (1978) does not relate solely to Neolithic migration (Zvelebil, 1989, 1998; Richards *et al*. 2003), given that, as we have noted, there have been other population movements in Europe from the Near East, such as the first colonisation of Europe by anatomically modern humans in the Upper Palaeolithic.

As in the case of classic markers, certain autosomal and Y-chromosome markers have likewise described

* Corresponding Author: Dr. Concepcion de la Rúa. Tel: 946 012 544; Fax: 946 013 500, E-mail: ggpruvac@lg.ehu.es

southeast-northwest gradients, in line with the hypothesis of Neolithic demic expansion (Chikhi *et al.* 1998a, 1998b; Rosser *et al.* 2000; Semino *et al.* 2000, 2004); Furthermore, Semino *et al.* (2004) have explained the variability of the Y-chromosome for certain markers with regard to more recent migrations. Likewise, when analysing the variability in a marker on the X-chromosome (*dys44*), Xiao *et al.* (2004) have observed a clinal distribution in the Eurasian continent, which disappears if only the European continent is considered. A possible explanation for this lies in the fact that the existence of a high rate of recent (post-Neolithic) female migration in Europe may have erased any previously existing gradient.

The first studies on the variability of the mitochondrial genome in Europe revealed a strong similarity amongst European populations, and an almost total absence of geographic structuring in the continent (Pult *et al.* 1994; Bertranpetit *et al.* 1995; Sykes *et al.* 1996; Richards *et al.* 1996; Simoni *et al.* 2000; Plaza *et al.* 2003). In addition, analysis of the variability of the mtDNA haplogroups in European and Near Eastern populations highlighted the existence of a SE-NW gradient (first component of the PCA, 51% of the variance) (Richards *et al.* 2002). According to the haplogroups that help to explain the first component, the SE-NW gradient observed is due to Palaeolithic migrations and, to a lesser extent, also to post-Neolithic migrations, so Richards *et al.* concluded that Neolithic farmers did not have a major influence on the genetic pool existing in Europe (Richards *et al.* 2002).

It has been suggested that the current Basque population constitutes a remaining vestige of European ancient Upper Palaeolithic populations (Bertranpetit *et al.* 1995). However, Richards *et al.* (1996, 2000) have suggested that the Basque population is considered an outlier for the majority of classic markers (Cavalli-Sforza *et al.* 1994) not because it is a vestige of pre-Neolithic populations, but due to a long period of isolation during which the differences in genetic drift would have been accentuated; likewise, this isolation would have attenuated the Neolithic genetic influence. Nonetheless, the Basque population has been used in various studies to represent the first European settlers of the Upper Palaeolithic (Wilson *et al.* 2001; Chikhi *et al.* 2002).

Up until now, the reconstruction of the biological history of European populations has been based on an analysis of the genetic variability of current populations, as ancient DNA (aDNA) data are very scarce at the population level. However, aDNA data are of great interest as they introduce a temporal factor into this historical reconstruction. These data are still insufficient at present, with few publications dealing with Europe's prehistoric, but anatomically modern, humans: the iceman found in the Alps (5,000 YBP) (Handt *et al.* 1994) and a further three individuals discovered in the Alps: Mezzocorona (6,400 YBP), Villabruna (14,000 YBP) and Borgo Nuovo (6,000 YBP) (di Benedetto *et al.* 2000); two individuals on the Paglici site (Southern Italy) dating back to around 24,000 years (Caramelli *et al.* 2003); 28 individuals belonging to the Etruscan population that populated Italy between the VII and III centuries B.C. (Vernesi *et al.* 2004); and 121 individuals from pre-historic times (3,500–5,000 YBP) in the Basque Country that have been previously analysed by our team (Izagirre & de la Rúa, 1999).

The present paper provides new genetic data on ancient populations settled in the region known today as the Basque Country, in order to assess the hypotheses proposed for the basis of present-day population data, regarding the origin of the genetic differentiation of the Basques. Therefore, we have analysed the mitochondrial genome of a historical population form the Basque Country (VI–VII c. A.D.), which we interpret together with the genetic data of various prehistoric populations (4,500–5,000 YBP approximately) that we have studied previoulsy (Izagirre & de la Rúa, 1999), in order to introduce an objective temporal factor into the analysis of the mtDNA variability of European populations. This analysis is based on the frequencies of mtDNA haplogroups, as they are the only data existing for Basque pre-historic samples.

Material and Methods

Material

The present work has analysed 76 individuals from a historical period (VI–VII c. A.D.), recovered from the necropolis of Aldaieta (Nanclares de Gamboa, Araba, Basque Country) (Table 1 and Fig. 1). The

Table 1 Archaeological Sites: location, chronology and sample size

Site	Location	Chronology	sample size
Aldaieta	Araba	VI–VII c. AD	76
SJAPL[a,b]	Araba	5,070 ± 150 – 5,020 ± 140 YBP	63
Longar[b]	Nafarroa	4,450 ± 70 – 4,580 ± 90 YBP	29
Pico Ramos[b]	Bizkaia	4,100 ± 110 – 4,790 ± 110 YBP	24

Sites in the Basque Country analysed by means of RFLPs of mtDNA: prehistoric (SJAPL, Longar and Pico Ramos) and historic (Aldaieta) populations. [a]: San Juan Ante Portam Latinam. [b]: Izagirre & de la Rúa (1999).

Figure 1 Geographic location of the ancient populations of the Basque Country considered in this work: the historical site of Aldaieta analysed in this paper and 3 prehistoric populations included in the analysis (SJAPL, Longar and Pico Ramos).

importance of this necropolis is because it is the only one providing human remains from the Late Antiquity period (Azkarate, 2004). Aldaieta was excavated by A. Azkarate's team between 1990 and 1993, following rigorous archaeological methods that enabled skeletal remains to be recovered, along with a great many everyday material assemblages that go along with human remains. A detailed record of the layout of the burials was made (Azkarate, 1999). Differences were observed in the distribution of material assemblages: some burials were extremely rich and included ornamental objects and weapons of various types (axes, knives, spearheads, etc.), while in others only coffin nails were found. Heterogeneity was also observed in the manner of burial (individual, in rows, and in groups of superimposed layers).

The comparative analysis has involved data corresponding to three prehistoric populations, analysed by our team in a previous paper (SJAPL, Longar and Pico Ramos) (Izagirre & de la Rúa, 1999, 2000) (Table 1 and Fig. 1), together with those existing in the database of Richards *et al*. (2000) which contains present-day populations from the Near East (1,234 samples), Europe (2,804 samples) and the northern part of the Caucasus (208) (Fig. 2). We have slightly restructured the composition of certain European regions considered in the database of Richards *et al*. (2000); an example is the population of Galicia which, due to its location in the extreme northwest of the Iberian Peninsula, has been excluded from the Western Mediterranean region defined by Richards *et al*. (2000), and is considered as an independent group in the present paper. This same approach has been adopted with the sample of Basques, which in the Richards *et al*. (2000) database was the only sample constituting Europe's southwest region (Fig. 2). In addition, we have also considered the T and X haplogroups jointly, because they were not differentiated in prehistoric populations. The genetic information on

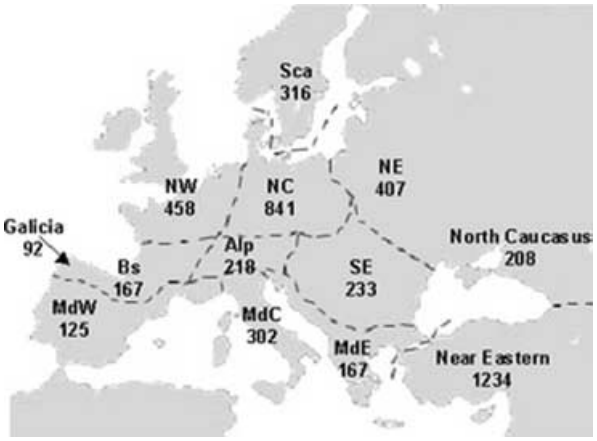

Figure 2 Samples of present-day European populations used for the comparative analysis grouped by geographical regions (modification of Richards *et al*. 2000) The geographic groups of Richard *et al*. (2000) have been maintained with the exception of the MdW region (from which the population of Galicia has been removed), and the SW region which we have renamed as Basques. The abbreviations are as follows: SE: Southeast; MdC: Central Mediterranean; MdE: Eastern Mediterranean; MdW: Western Mediterranean; NE: Northeast; NC: Northern-Central Europe; Alp: Alps; NW: Northwest; Sca: Scandinavia; Bs: Basques. The number below the name of the populations indicates the size of the sample.

the present-day Basque population has been completed with data taken from the publication by Larruga *et al*. (2001).

Methods

The genetic analysis of the DNA recovered from the archaeological remains involves a number of limitations, namely the scarcity and fragmentation of DNA that is recovered, and the risk of contamination. In this case the entire excavation process involved strict precautions designed to avoid contamination. Furthermore, the anthropological remains were immediately removed to our laboratory, without undergoing washing, and often still embedded in the sediment, where cleaning was performed in dry conditions. The samples used in this paper are dental pieces, the most isolated system in skeletal remains and therefore less liable to outside contamination. The processing of the samples in the laboratory involved the application of a series of strict criteria detailed in Cooper & Poinar (2000) and Hofreiter *et al*. (2001) for the authentication of results. In our case, the extraction and preparation of the PCR was undertaken in a positive-pressure sterile chamber, physically separated from the laboratory where the post-PCR processes are carried out. All the work surfaces were cleaned regularly with sodium hypochloride and irradiated with UV light. Suitable disposable clothing was worn (lab coat, mask, gloves and cap). Contamination controls were applied in both the extraction and amplification processes. A duplicate analysis was performed for all individuals, insofar as possible, and some samples were analysed in an independent laboratory (INT, Madrid). Quantification of target DNA was carried out in a sub-sample by real-time PCR (Alonso *et al*. 2004). The confirmation of the haplogroups obtained by means of the PCR-RFLP methodology involved the sequencing and cloning of HVR I of the mtDNA in 10 individuals. Finally, the results obtained for the ancient sample were compared with the haplogroups and sequences obtained for the researchers who handled the samples.

Collection, Selection and Preparation of Dental Samples

Selection was made of those teeth that do not have caries or deep fissures that might extend into the pulp. Whenever possible, more than one tooth was taken from each individual for duplicate analysis, proceeding to analyse the duplicates in different sessions. In addition, 9 samples were analysed in an independent laboratory.

Extraction of DNA Using Phenol/Chloroform

In order to eliminate surface contamination, the teeth were subjected to a process of depurination using acids, and the entire surface was irradiated with ultraviolet light (Ginther *et al*. 1992). The extraction process followed the protocol described by Hagelberg & Clegg (1991): after cutting the root of the tooth, it was incubated with stirring over night at 37° C, in a lysis buffer (5 ml) (0.5M EDTA pH 8.0–8.5; 0.5% SDS; 50mM Tris HCl pH 8.0; 0.01mg/ml proteinase K). The recovery of the DNA involved the use of phenol and chloroform and was finally concentrated and purified by means of columns (Centricon-30, Amicon). Each extraction session involved 2 contamination controls that were applied to the entire extraction process, with the difference being that no dental tissue was added. Three proportional parts

Table 2 Primer sequences, annealing temperature and product length

Marker	Primer sequence	Annealing temperature (°C)	Amplicon length
Mse I	L-tcaactacaagaacaccaatgacc	52°	125pb
14766	H-agtgagccgaagtttcatcatg		
Dde I	L-cttattaatcatcatcctagc	54°	90pb
10394	H-ttgtttaaactatataccaattcgg		
Alu I	L-accgtaggtggcctgactgg	62°	120pb
7025	H-ggcaaatacagctcctattgataggac		
Nla III	L-cactcatcacagcgctaagc	55°	120pb
4577	H-tggcagcttctgtggaac		
Nla III	L-aacttcctaccactcacc	47°	121pb
4216	H-tactctatcaaagtaactct		
Hae II	L-cctaaccgctaacattac	51°	120pb
9052	H-gaagatgataagtgtagagg		
Hinf I	L-cacaagaactgctaactcatgc	55°	123pb
12308	H-cttttatttggagttgcaccaagatt		
Dde I	L-taacttgaccgctctgagct	57°	102pb
1715	H-cttgcggtactatatctattgc		
Alu I	L-ttgatgagggtcttactc	46°	118pb
10032	H-tagtagtaaggctaggagg	(Hot-Start)	
Acc I	L-caccaagacctcaacccctg	60°	95pb
14465	H-atttaggggggaatgatggttgtc		
Hae III	L-ttcttaccacaaggcacacc	65°	126pb
8994	H-aggtggcctgcagtaatgt		

Sequences of primers designed for the amplification of the ancient mtDNA, annealing temperature of each and size of the amplified product.

of each DNA extract were made in order to overcome the drawbacks due to possible contamination.

Analysis of the RFLPs of the Coding Region of the Mitochondrial DNA

In order to classify the variability of the mitochondrial genome of the individuals buried in Aldaieta we proceeded to amplify and type 11 markers, which are required for defining the 10 haplogroups that have been described in Caucasoids (Macaulay *et al*. 1999).

The amplification involved the design of a pair of primers for each marker that amplify a product of 100–120 bp in length, approximately. Table 2 shows the list of *primers*, the annealing temperature of each one and the size of the amplified product. Amplifications were performed in a 35 μl reaction mixture containing Bioline buffer 1X, 1mM of $MgCl_2$, 0.1 μM of each dNTP, 0.02 mg/ml of BSA, 0.3 μM of each primer, 1.5 unit of *Taq* polymerase (Bioline) and 10 μl of diluted ancient DNA (1:15). Cycling parameters were 94°C for 5 min, followed by 40 cycles of 94°C for 30 sec, annealing temperature for 10 sec., 72°C for 30 sec, and 72°C for 10 min. Apart from the two extraction controls corresponding to each sample, each PCR included a contamination control of the PCR, which was applied to the entire amplification process but in which the DNA extract was replaced by water. The results of the PCR were verified in agarose gels dyed with ethidium bromide. If both the PCR controls and the extraction blanks were contamination free, the positive samples were subjected to enzymatic digestion as per the suppliers' instructions. The absence/presence of digestion was verified in 12% acrylamide gels and with silver-staining. Table 3 shows the restriction map which was used to define the mtDNA haplogroups that have been described in Caucasoids.

Authentication Methods

In addition to the precautions taken to avoid contamination, application was made of other authentication criteria such as quantification, sequencing and cloning:

Quantification of target DNA: The quantification of amplifiable DNA was performed by means of real-time PCR of two fragments, one of 113bp and the other

Table 3 mtDNA haplogroups restriction pattern

	*Mse*I 14766	*Dde*I 10394	*Alu*I 7025	*Nla*III 4577	*Nla*III 4216	*Hae*II 9052	*Hin*fI 12308	*Dde*I 1715	*Alu*I 10032	*Acc*I 14465	*Hae*III 8994
H	−	−	−								
V	−	−	+	−							
HV	−	−	+	+							
T		−			+						
J		+			+						
U		−				+	+				
K						−	+				
I								−	+		
X								−		+	
W											−

The restriction pattern with absence (−)/presence (+) of the enzymatic digestion corresponding to the markers analysed in the present study for determining the mtDNA haplogroup of Caucasoids (Macaulay *et al*. 1999).

of 287bp, using Taqman probes in a sub-sample of 9 individuals. The methodology is described in Alonso *et al*. (2004).

mtDNA HVR I sequencing: The sequencing of the HVR I (nps 15,998–16,400 as per Anderson, 1981) was undertaken for 56 individuals in 6 overlapping fragments of approximately 100 bp each. Application was made of the methodology described by Alonso *et al*. (2003), although the amplification of each fragment was undertaken in independent PCRs rather than in two multiplex-PCRs. The PCRs were performed in 50 μl of reaction mixture containing GeneAmp Buffer 1X, 2 mM of $MgCl_2$, 0.1 μM of each dNTP, 0.4 μM of each primer, 5 units of AmpliTaq Gold (Applied Biosystems) and 20 μl of diluted DNA (2 μl of DNA extract in 18 μl of 1mg/ml BSA). Cycling parameter were 95°C for 10 min; followed by 36 cycles of 95°C for 10 sec, annealing temperature for 30 sec, 72°C for 30 sec; and 62°C for 10 min. The annealing temperatures of the primers were as follows: 60°C for the A1/A1R primer pair, 58°C for 2F/2R and 4F/4R, 57°C for 1F/1R and 55°C for 3F/3R and 5F/5R (the primer sequences are listed in Alonso *et al*. 2003). In the event of positive amplification and absence of contamination, the amplifications were purified by means of Centricon-100 (Amicon) and subsequent sequencing in an ABI310 automatic sequencer using chemistry based on dRhodamine. The results obtained were edited with the BioEdit software application and the sequences aligned manually (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Replication: The sequencing of HVR I of a sub-sample of 9 individuals was carried out at the Instituto Nacional de Toxicología (INT) in Madrid through the application of the methodology described by Alonso *et al*. (2003). The entire process (extraction and analysis of DNA) was undertaken by different researchers.

Cloning: In order to detect possible heterogeneities in the PCR products (not visible with direct sequencing) that may correspond to either post-mortem damage and/or mixed contamination, cloning was undertaken of a fragment of the HVR I of a sub-sample of 10 individuals by means of the TOPO TA Cloning® Kit (Invitrogen). Linkage to the vector pCR®2.1-TOPO® and chemical transformation of the cells TOP10F' (One Shot® *E. coli*) were performed by following the supplier's instructions. After culturing the cells overnight at 37°C in a selective environment, 8 blank cultures were collected and each cultured in a non-selective environment. Five μl of purified product (by means of QIAprep Spin Mini prep Kit, QIAGEN) of each clone were used for the sequencing reaction.

Statistical Analysis

We carried out a principal components (PC) analysis in which the frequencies of the mtDNA haplogroups found in this study on the historical population of Aldaieta were considered, together with the frequencies obtained from several prehistoric sites in the Basque Country that we have studied so far (SJAPL, Longar and Pico Ramos) (Izagirre & de la Rúa, 1999). These data on past populations were analysed within the

context of present-day populations in the Near East and Europe (4,246 samples in the database of Richards *et al*. 2000; Larruga *et al*. 2001).

A distance matrix was estimated (Reynolds *et al*. 1983) using the GENDIST program included in PHYLIP (Felsenstein, 1989) with the frequency data of the mtDNA haplogroups of the 22 populations considered in the PC analysis. The resulting genetic distance matrix was represented in two dimensions by the use of multidimensional scaling analysis (Kruskal, 1964), implemented by the ALSCAL program included in the SPSS package, version 12.0.

Results

In this paper we have analysed the variability of the mitochondrial genome by means of RFLP of 76 individuals recovered at the historical site of Aldaieta (VI–VII c. AD) (Basque Country). The primers designed give an amplified product that is around 100–120 bp, suitable for the analysis of degraded samples, which enabled us to determine the mtDNA haplogroup of 67 individuals (80% amplification efficiency) and the HVR I sequences of 48 individuals. Additional information regarding the samples is available as supplementary material.

Statistical Considerations: Correction of the Frequencies of the Haplogroups

Some of the results forthcoming in the analysis of the mtDNA lead us to suspect the presence of kinship on the maternal side in the Aldaieta necropolis. Considering the HVR I of 48 individuals that were sequenced, we defined 16 different haplotypes. Some of the haplotypes described, which are no longer common today, belong to individuals interred in the same burial group (for example, 8 of the 9 individuals who share the same haplotype are found in a group burial consisting of 13 individuals). Furthermore, the study of specific mutations such as nt73 of the HVR II, and archaeological data, also support the existence of kinship (Izagirre *et al*. 2005 - under revision). The criterion used for the correction of the frequencies obtained, similar to that described by Vernesi *et al*. (2004), is that when a haplotype is represented by two or more individuals interred in the same burial group, only one of them has been considered. Taking into account this criterion, the haplogroup frequencies of the population of Aldaieta have been corrected, whereby only 37 individuals were considered for subsequent statistical analyses. The corrected frequencies of the haplogroups of the historical population of Aldaieta, together with the frequencies of the pre-historic and present-day Basque populations, are displayed in Table 4.

Authentication of Results

The amount of amplifiable DNA in a sub-sample of 9 individuals were quantified for one fragment of 113 bp and another of 287 bp by real-time PCR. All the samples gave positive results for the smaller fragment, the majority (8 out of 9) with between 50–500 copies/μl and one with 3,000 copies/μl. This latter sample is the one that gave a positive result for the 287 bp fragment, recording less than 50 copies per microlitre (Alonso *et al*. 2004; this information is also available as supplementary material).

Table 4 MtDNA haplogroup frequencies

	Aldaieta frec ± s.d.	SJAPL[a] frec ± s.d.	Longar[a] frec ± s.d.	Pico Ramos[a] frec ± s.d.	present-day Basques[b] frec ± s.d.
H,HV	0.486 ± 0.0833	0.377 ± 0.0626	0.440 ± 0.1013	0.375 ± 0.1009	0.623 ± 0.0376
J	0.162 ± 0.0614	0.164 ± 0.0478	–	0.167 ± 0.0777	0.024 ± 0.0119
V	0.027 ± 0.0270	–	–	–	0.102 ± 0.0235
T,X	0.108 ± 0.0518	0.049 ± 0.0279	0.160 ± 0.0748	0.167 ± 0.0779	0.060 ± 0.0184
U	0.162 ± 0.0614	0.180 ± 0.0496	0.160 ± 0.0748	0.125 ± 0.0690	0.138 ± 0.0267
W	–	–	–	–	0.012 ± 0.0084
K	0.027 ± 0.0270	0.229 ± 0.0543	0.240 ± 0.0872	0.167 ± 0.0777	0.036 ± 0.0144
I	0.027 ± 0.0271	–	–	–	–

Distribution of the corrected frequencies of the mtDNA haplogroups of the historical population of Aldaieta (VI–VII c. AD) analysed in this study. Frequencies of three prehistoric populations in the Basque Country (SJAPL, Longar and Pico Ramos) and those of the present-day Basque population used for comparison. ([a]) data by Izagirre & de la Rúa (1999); ([b]) data taken from Bertranpetit *et al*. (1995), Corte-Real *et al*. (1996), Torroni *et al*. (1998), Richard *et al*. (2000), Larruga *et al*. (2001). frec.: frequencies; s.d.: standard deviation.

Of the 37 individuals from Aldaieta considered in the statistical analysis, after correcting the frequencies to avoid kinship 94.6% were analysed in duplicate, with the results coinciding in all cases. 34 individuals (91.9%) were sequenced, 25 of them in duplicate, and 8 samples (21.6%) were replicated in an independent laboratory (INT, Madrid). The sequencing analysis of the HVR I confirmed the results obtained by means of RFLPs (Table 5) as, on one hand the motifs described by Macaulay *et al*. (1999) for the HVR I coincided with the determination of haplogroups carried out by means of RFLP (criterion of authentication proposed by Montiel *et al*. 2001) and, on the other, wide variability was found in the HVR I sequences. A further authentication criterion was cloning 10 previously sequenced individuals. Between 5 and 8 clones were obtained for a fragment of the HVR I from each individual, and in all cases the consensus sequence coincided with the fragment analysed by direct sequencing (Fig. 3).

Taking into account that (1) the HVR I motifs of the mtDNA of the sequenced individuals coincided with the haplogroup determined by RFLPs; (2) 29% of the sequences were authenticated by means of cloning; and (3) application was made of other common authentication criteria in the analysis of aDNA (see material and methods), we deduce that the present results are not artefacts or the result of contamination.

Table 5 Number of individuals from Aldaieta following the correction of frequencies for kinship determined by RFLPs and by sequencing

Haplogroup	n_r	n_s	HVR I Motifs
H	18	18	–
J	6	5	069, 126
		2	051G, 129C, 362
U	6	3	270
		1	189, 249
T	3	3	126, 294
K	1	1	224, 311
V	1	1	298
I	1	–	–
X	1	–	–
TOTAL	37	34	

The number of individuals whose haplogroup has been determined by means of PCR-RFLP (n_r), by sequencing of HVR I (n_s) and the HVRI motifs each haplogroup presents. The individuals in this table are solely those considered for the statistical analysis following frequency correction for kinship.

Principal Component (PC) Analysis

Figure 4a presents the first two components of the multivariate (PC) analysis performed (63.1% of the overall variability). In this analysis we considered jointly the present-day populations in the Near East and Europe (database of Richards *et al*. 2000; Larruga *et al*., 2001), together with 3 prehistoric populations from the Basque Country (SJAPL, Longar and Pico Ramos) (Izagirre &

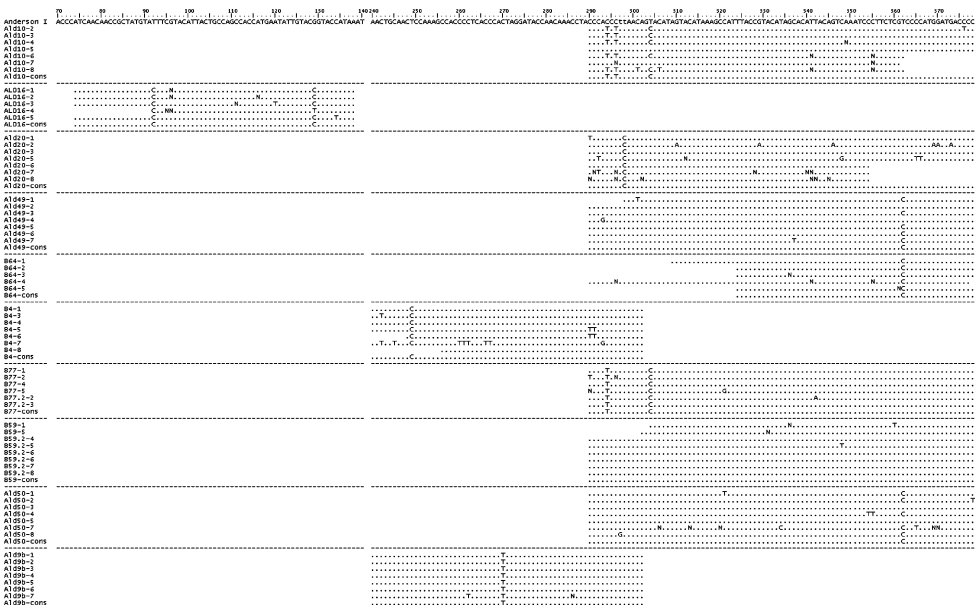

Figure 3 The sequencies of HVR I cloned after amplification from 10 individuals from Aldaieta. The last sequence of each individual is the consensus sequence. Dots indicate identity to the reference sequence (Anderson *et al*. 1981).

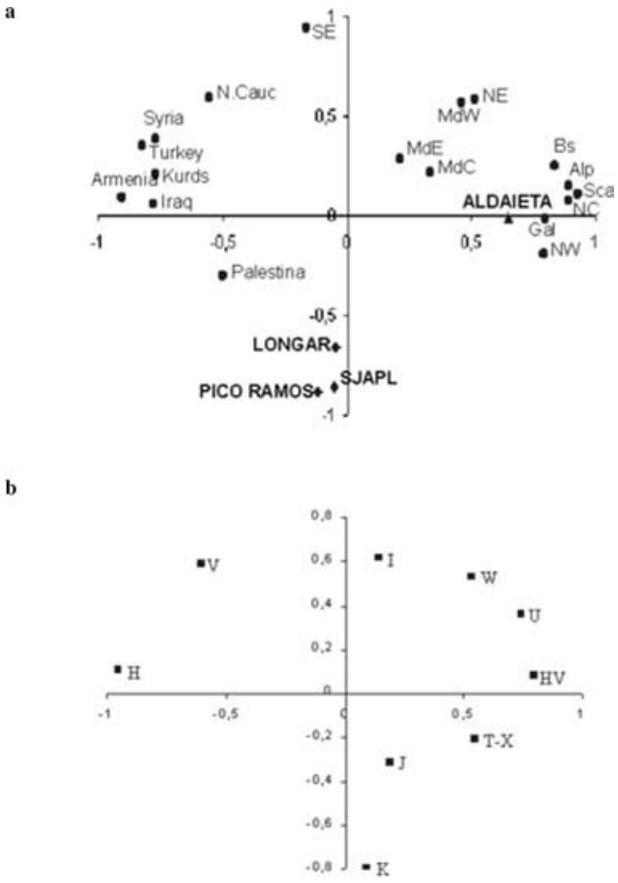

Figure 4 a) Analysis of the main components considering the frequencies of the mtDNA haplogroups of the historical population of Aldaieta (◊), three prehistoric Basque populations (SJAPL, Longar and Pico Ramos) (▲) and several present-day populations in the Near East and Europe (●). Gal: Galicia; the remaining abbreviations are explained in Figure 2. b) Correlation of the mtDNA haplogroups with the first two components of the analysis.

de la Rúa, 1999) and the historical population of Aldaieta analysed in this paper. The first component, which accounts for 42.6% of the total variance, established a differentiation between the present-day populations of the Near East and those of Europe. Within Europe, the populations of the Mediterranean area (MdE, MdC and MdW) and those of Eastern Europe (NE and SE) are closer to those of the Near East. Regarding the prehistoric populations of the Basque Country, they are situated between the two groups (Europe and Near East), whereas the historical population of Aldaieta falls within the variability range of present-day European populations. The main haplogroups that explain the variability registered by the first component are H, HV, V and U (Fig. 4b). The second component, which accounts for 20.5% of the overall variance, sets the prehistoric populations apart from both present-day populations as well as the historical one of Aldaieta (Fig. 4a). The principal haplogroups involved in this differentiation are haplogroup K and, to a lesser extent, I, V and W (Fig. 4b).

A second PC analysis was performed in which only European populations were considered (Fig. 5). The first two major components of this analysis account for 54.7% of the overall variance. In the first component (which accounts for 32.1% of the variance) we observe a separation of the prehistoric populations of the Basque Country from both present-day populations as well as from the historical one of Aldaieta (Fig. 5a). Haplogroup K explains a large part of the variability recorded in the first component and, likewise, haplogroups W and V (Fig. 5b). The second component (which accounts for

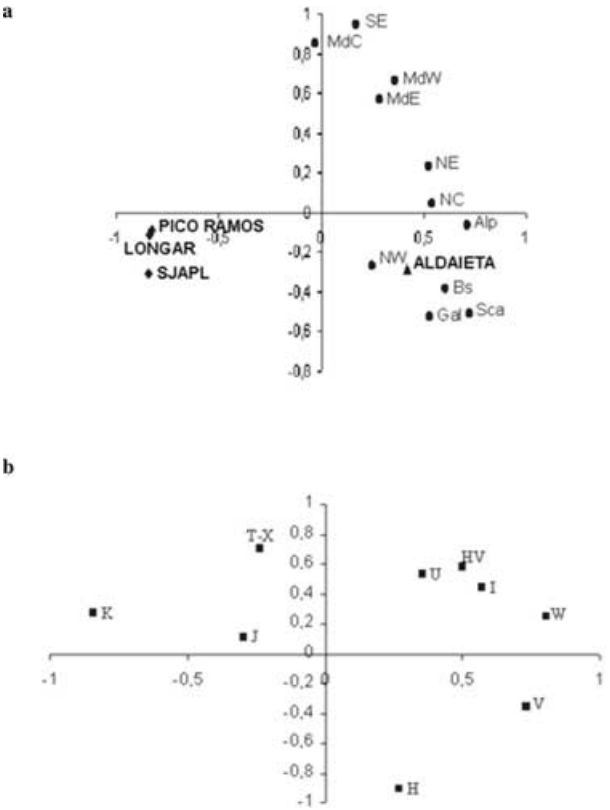

Figure 5 a) Analysis of the main components considering the frequencies of the mtDNA haplogroups of the historical population of Aldaieta (▲), three prehistoric populations in the Basque Country (Pico Ramos, Longar and SJAPL) (⧫), and several present-day European populations (●). Gal: Galicia; the remaining abbreviations are explained in Figure 2. b) Contribution of the mtDNA haplogroups to the first two components of the analysis performed.

22.6% of the total variance) shows present-day European populations distributed into several groups: one of these, which we shall refer to as the Mediterranean area, consists of the regions around the Mediterranean basin (SE, MdC, MdW, MdE); another contains the populations located along the Atlantic fringe (NW, Bs, Sca and Gal) together with the historical population of Aldaieta; and a third group is made up of the populations of Central Europe (NE, NC, Alp), which is situated between the two preceding groups. Haplogroup H, together with the sum of haplogroups T-X, accounts for a large part of the variability recorded by this second component (Fig. 5b)

The *Western Mediterranean* region (MdW), as defined in the database of Richards *et al*. (2000), is situated close to the regions of Central and Northern Europe, being differentiated from the other Mediterranean regions. However, in this paper, having excluded the population of Galicia from the MdW, we note the proximity of MdW to the regions that constitute the Mediterranean area and the location of Galicia within those of the Atlantic fringe (Fig. 5a).

Haplogroup contribution to PCA.

As we have seen, haplogroup K has a considerable bearing on the distribution of modern and ancient populations in both PC analyses. This haplogroup is at a high frequency in the prehistoric populations of the Basque Country (16% in Pico Ramos and around 23% in Longar and SJAPL) (Table 4), whereas the average value in present-day European populations is 4.8%, ranging between 3.6% and 7.7% (Richards *et al*. 2000).

Haplogroup V has a bearing on both components 1 and 2 of the analysis undertaken considering the populations of the Near East (Fig. 4), by being absent in the majority of these Eastern populations and in all the prehistoric ones from the Basque Country. For the latter reason it also has an influence on the first component of the analysis performed considering solely European populations (Fig. 5). The average frequency of haplogroup V in the present-day population of Europe is 4.8%, one of its highest frequencies being in the present-day Basque population (average value: 10.2% as per Larruga *et al*. 2001). Torroni *et al*. (1998) suggested that the origin of haplogroup V lies in Southwest Europe, and from there it spread to other European regions between 10,000–15,000 YBP. As we have already seen, we did not find any individuals in the prehistoric populations that belong to lineage V (Izagirre & de la Rúa, 1999) and we found only one individual (2.7%) in the historical population of Aldaieta. On the other hand, Maca-Meyer *et al*. (2003) propose that the origin of haplogroup V should be displaced to Cantabria. Should this be the case, the presence of haplogroup V in the historical population of Aldaieta suggests the existence of genetic flow between the population of the Basque Country and that of the neighbouring region, at least since Late Antiquity.

As we have seen, haplogroup H is one of the haplogroups that determined the differentiation between the populations of the Near East and Europe (Fig. 4). When the PC analysis was carried out taking into account only European populations (Fig. 5), haplogroup H, together with the sum of haplogroups T-X, determined the differentiation between several population groups: Mediterranean area, Atlantic fringe and Central Europe (Fig. 5a). Haplogroup H is the one most represented in all the populations considered in this study, showing a lower frequency in the Near East, whilst its highest frequency is in the Southwest of Europe (Galicia 64% and Basque Country 62%). The frequency of haplogroup H in the historical population of Aldaieta (48.6%) is similar to that displayed in the present-day Atlantic fringe populations. However, the prehistoric populations of the Basque Country studied prior to this show a lower frequency of haplogroup H, as in Longar it is 44% and in the other two prehistoric populations (SJAPL and Pico Ramos) its frequency is around 37% (Table 4).

Haplogroup J has been the main lineage of mtDNA related to the Neolithic expansion from the Near East around 10,000 years ago (Richards *et al*. 1996, 2000). A lower frequency of J has been displayed in the present-day Basque population (2.4%), whereas in all other regions of Europe its frequency ranges between 7% and 14%. However, in certain prehistoric populations from the Basque Country (SJAPL and Pico Ramos), and in the historical population of Aldaieta, the presence of this haplogroup is greater than in the rest of the European populations studied, displaying values of around 16%, although it is absent in the prehistoric population of Longar (Table 4). Nevertheless, haplogroup J is of no relevance in either of the first two components in the PC analysis performed in this paper (Fig. 4b and 5b).

Multidimensional Scaling

The representation of the genetic distance matrix, based on haplogroup frequencies by means of multidimensional scaling, concurs with the result obtained through main component analysis, as it establishes a clear distinction between the prehistoric populations and the present-day and historical one from Aldaieta (Fig. 6). Within the present-day populations, those of the Near East are grouped on the opposite side to those of Europe. Within the European ones, a similar distribution is observed to that noted in the PC analysis, establishing a differentiation between the populations in the Mediterranean area and those of the Atlantic fringe. The historical population of Aldaieta lies close to the populations of the Atlantic fringe. The two-dimensions account for 96.1% of the total variation, with the stress value of 9.8% indicating a good fit between the graph and the original distance matrix.

Discussion

The frequency of haplogroup H is an outstanding feature in the multivariate analyses and establishes a differentiation between populations of the Near East and Europe (Fig. 4), and within the Europeans it differentiates the regions of the Mediterranean basin and those of the Atlantic fringe (Fig. 5). Haplogroup K also accounts for a large part of the variance recorded in both PC analyses. This is due to the frequency variation between the values in the present-day European populations (4.8%) and the historical one (2.7%) on the one hand, and on the other, the high values for the prehistoric populations of the Basque Country (16%–23%), in which it is the most common haplogroup after H (Table 4). The historical population of Aldaieta falls within the variability displayed by the regions of the Atlantic fringe. Nevertheless, the three prehistoric populations of the Basque Country stand apart from those of the Atlantic fringe (Fig. 4, 5 and 6).

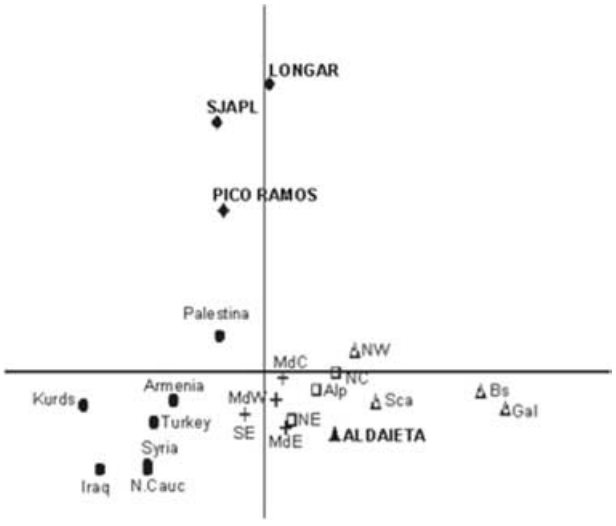

Figure 6 Multidimensional scaling analysis for 22 populations, based on a genetic distance matrix of present-day Near Eastern (●) and European populations (Mediterranean area (+); central Europe (□), Atlantic Fringe (△), three prehistoric populations from the Basque Country (Pico Ramos, Longar and SJAPL) (⧫) and the historical one from Aldaieta (▲).

When comparing the frequencies of the mtDNA haplogroups, we observed statistically significant differences between the present-day Basque population and each one of the three prehistoric ones for the haplogroups as a whole ($p < .001$), and likewise for haplogroups K ($p < .005$), H ($p < .01$) and V ($p < .001$), for the high values of K, lower ones for H and absence of V in the prehistoric ones. This might be due to endogenous factors in the prehistoric samples, such as the existence of a certain degree of kinship amongst the sample individuals, which would have produced an increase in the frequency of certain haplogroups (in this case of hg K). However, this is fairly unlikely given the coincidence that the three prehistoric populations show a high frequency of the same haplogroup (K). It may also be considered that these prehistoric populations from the Basque Country were not indigenous, but rather groups linked to Neolithic farmers from other regions. In the case of the sites of SJAPL and Longar this might be a plausible hypothesis, as they are located in the Ebro Valley, which is one of the routes Neolithic culture followed on its way into the Iberian Peninsula. However, Pico Ramos is a site located on the Atlantic coastline of the Basque Country, whose dwellers were perfectly adapted to the setting, living on the resources provided by the nearby marshlands (Baraybar & de la Rúa, 1997). Both the anthropological analysis and the archaeological data consider Pico Ramos as an indigenous settlement of the Chalcolithic period that had a Neolithic culture. Given that the three prehistoric populations are grouped within the PC analysis and in the multidimensional scaling plot carried out, we assume that all three (SJAPL, Longar and Pico Ramos) are part of a common genetic

pool. If this is the case, the differences observed in the frequencies of the haplogroups of the prehistoric populations and the present-day one may be attributed to evolutionary processes experienced by the Basque population some time after the Neolithic.

One of the explanations for the differences between the prehistoric and present-day populations may be natural selection, as there may be adaptive advantages of certain Caucasian haplogroups of mtDNA over others (Mishmar *et al*. 2003; Wallace *et al*. 2003; Ruiz-Pesini *et al*. 2004). In the light of present-day knowledge, this does not seem to be a plausible interpretation as, on the one hand, although a link has been established between specific haplogroups and certain illnesses (for example, haplogroups K and J appear to be linked to multiple sclerosis), protection has also been detected against other illnesses (hg K and J protect against Alzheimer's and Parkinson's). Another possible interpretation for the temporal differentiation of the frequencies of the mtDNA haplogroups is the genetic drift brought about by a bottleneck, which above all would have led to a decrease in lineages of haplogroup K and an increase in the frequency of haplogroups H and V. Finally, the change in frequencies might be the result of gene flow. We cannot dismiss female gene flow between populations on the Atlantic fringe, as these populations show similarities amongst each other. A greater differentiation has been detected between European populations in terms of Y-chromosome and autosomal markers than mtDNA, which has suggested the existence of greater female mobility (Seielstad *et al*. 1998). Furthermore, certain studies have detected a post-Neolithic gene flow in Europe (Semino *et al*. 2004; Xiao *et al*. 2004) which, in the case of markers related to female lineages (mtDNA and markers of the X-chromosome), may have erased the possible gradients produced by both Palaeolithic and Neolithic expansion.

Like us, Vernesi *et al*. (2004) detected a significant difference between the present-day population of Tuscany and the prehistoric one inhabiting the same area (Etruscan), with a gap of 2,500 years between them, finding only two haplotypes in common between both populations. It may be that this phenomenon is more widespread and has occurred in other regions of Western Europe. Data on more prehistoric populations are required in order to confirm this phenomenon. It should also be taken into account that these differences have only been detected in the mtDNA, as currently, the nuclear genome of prehistoric European individuals has not been studied at the population level.

In terms of drawing inferences based solely on data from present-day populations, all the changes that may have occurred in the prehistory and history of Europe should be taken into account, and as all prehistoric and numerous historical events lack written documentation, the importance of the DNA data recovered from skeletal remains is considerable, and may be illustrated by several results from this study. Given that the present-day Basque population, considered a vestige of the first Palaeolithic populations (Bertranpetit *et al*. 1995, amongst others), has one of the highest frequencies of H shown in Europe, certain authors point out that it might be the characteristic of European settlers of the Upper Palaeolithic (Gonzalez *et al*. 2003). Thus, they suggest that the geographic gradient displayed by haplogroup H may be due to differences in the impact of subsequent migrations on a common Palaeolithic substratum. However, as we have already seen, the prehistoric populations of the Basque Country studied so far have a lower frequency of H. Bearing in mind this temporal variation in the frequency of haplogroup H in the Basque Country, it is not correct to state that the widespread situation in the Upper Palaeolithic involved the existence of high frequencies of haplogroup H, simply due to the fact that present-day Basques present this characteristic. Another example is the case of haplogroup J. Given that the present-day Basque population is an outlier regarding the Neolithic component, it has been proposed that this region experienced a smaller genetic impact from Neolithic farmers. But if we accept that lineage J is a marker of migrations of Neolithic populations from the Near East, then the Basque Country also experienced the impact of these peoples, as is shown by the high frequency of haplogroup J in certain ancient populations. In addition, the heterogeneity this haplogroup shows in different prehistoric groups might suggest that the adoption of Neolithic culture followed different paths within the same region. These data indicate that certain inferences based solely on the frequencies in present-day populations do not appear to be correct. As advised by Vernesi *et al*. (2004), this leads us to reconsider the supposition whereby the genetic patterns of present-day

populations reflect the evolutionary processes experienced by their predecessors (Sokal *et al*. 1991; Richards *et al*. 2000, 2002, amongst others). Up until now it was thought that there was no evolutionary process subsequent to the Neolithic that altered the genetic composition of European populations (Barbujani & Bertorelle, 2001). However, our data on ancient DNA (as well as those of Vernesi *et al*. 2004) reveal a discontinuity between prehistoric and present-day populations, which leads us to reconsider the limitations involved in the reconstruction of evolutionary history on the basis of the genetic patterns of present-day populations.

In conclusion, taking into account the differences found between the past and present-day Basque populations, we suggest that the distribution of the frequencies of the mtDNA haplogroups in present-day populations may have been due to a post-Neolithic restructuring of the population (the prehistoric sites studied date from the end of the Neolithic-Chalcolithic). Bearing in mind that the historical population of Aldaieta (VI–VII c. AD) shows frequencies that are included in the variability of the present-day populations of the Atlantic fringe, we may assume that this restructuring had already taken place in Late Antiquity. In short, between 5,000–1,500 YBP there was a restructuring of the distribution of the frequencies of mitochondrial DNA in the population of the Basque Country, and possibly in more regions of Western Europe, probably as a result of gene flow and/or genetic drift.

In order to provide a more specific explanation of the differences found in the frequencies of the mtDNA haplogroups amongst the populations of the Basque Country, both ancient and present, analysis is required of a greater number of populations dating from between 5,000 and 1,500 YBP in order to interpret the meaning and rate of these changes.

Furthermore, the present study on mtDNA haplogroups points to the existence of a recent relationship amongst the historical population of Aldaieta and those constituting the Atlantic fringe. The nature of this relationship is being studied in greater depth by the analysis of the sequences of the control region of mtDNA.

Acknowledgements

We should especially like to thank the three anonymous reviewers for their helpful suggestions. This work was supported by funds from the University of Basque Country (1/UPV/EHU 00154.310-EA-8130/2000 and 9/UPV 00154.310–14495/2002) and from the Spanish Ministry of Science and Technology (MCyT) (BOS2000–0408).

References

Alonso, A., Albarrán, C., Martín, P., García, P., García, O., de la Rúa, C., Alzualde, A., Fernández de Simón, L., Sancho, M. & Fernández-Piqueras, J. (2003) Multiplex-PCR of short amplicons for mtDNA sequencing from ancient DNA. *International Congress Series* **1239**, 585–588.

Alonso, A., Martín, P., Albarrán, C., García, P., García, O., Fernández de Simón, L., García-Hirschfeld, J., Sancho, M., de la Rúa, C. & Fernández-Piqueras, J. (2004) Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Science International* **139**, 141–149.

Ammerman, A. & Cavalli-Sforza, L. L. (1973) A population model for the diffusion of early farming in Europe. In: *The Explanation of Culture Change*,C. Renfrew, Ed. (Duckworth, London).

Ammerman, A. & Cavalli-Sforza, L. L. (1984) *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton)

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Azkarate, A. (1999) Necrópolis tardoantigua de Aldaieta (Nanclares de Gamboa, Alava). Volumen I. *Memorias de yacimientos alaveses n° 6*, Diputación Foral de Álava, Departamento de Cultura. Vitoria/Gasteiz. 536 pp.

Azkarate, A. (2004) Reihengräberfelder al sur de los Pirineos occidentales?. *Sacralidad y arqueologia: Homenaje al Prof. Tuilo Ulbert,* (eds. J.M. Blazquez & A. Gonzalez), pp 389–413.

Baraybar, J. P. & de la Rúa, C. (1997) Reconstrustion of Diet with Trace Elements of Bone at the Calcolithic Site of Pico Ramos, Basque Country, Spain. *J Archaeol Sci* **24 (4)**, 355–364.

Barbujani, G. & Bertorelle, G. (2001) Genetics and the population history of Europe. *Proc Natl Acad Sci USA* **98**, 22–25.

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

Caramelli, D., Lalueza-Fox, C., Vernesi, C., Lari, M., Casoli, A., Mallegni, F., Chiarelli, B., Dupanloup, I., Bertranpetit, J., Barbujani, G. & Bertorelle, G. (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc Natl Acad Sci USA* **100**, 6593–6597.

Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton).

Chikhi, L., Destro-Bisol, G., Bertorelle, G., Pascali, V. & Barbujani, G. (1998a) Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc Natl Acad Sci USA* **98**, 9053–9058.

Chikhi, L., Destro-Bisol, G., Pascali, V., Baravelli, V., Dobosz, M. & Barbujani, G. (1998b) Clinal variation in the nuclear DNA of Europeans. *Hum Biol* **70**, 643–657.

Chikhi, L., Nichols, R. A., Barbujani, G. & Beaumont, M. A. (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci USA* **99**,11008–11013.

Cooper, A.& Poinar, H. N. (2000) Ancient DNA: Do it right or not at all. *Science* **289**, 1139.

Di Benedetto, G., Nasidze, I. S., Stenico, M., Nigro, L., Krings, M., Lanzinger, M., Vigilant, L., Stoneking, M., Pääbo, S. & Barbujani, G. (2000) Mitochondrial DNA sequences in prehistoric human remains from the Alps. *Eur J Hum Genet* **8**, 669–677.

Felsenstein, J. (1989) PHYLIP- Phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.

Ginther, C., Issle-Tarver, L. & King, M. C. (1992) Identifying individuals by sequencing mtDNA from teeth. *Nat Genet* **2**, 135–138.

Gonzalez, A. M., Brehm, A., Pérez, J. A., Maca-Meyer, N., Flores, C. & Cabrera, V. M. (2003) Mitochondrial DNA Affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol* **120**, 391–404.

Hagelberg, E. & Clegg, J. B. (1991) Isolation and characterization of DNA from archaeological bone. *ProcR Soc Lond B* **244**, 45–50.

Handt, O., Richards, M., Trommsdorff, M., Kilger, C., Simanainen, J., Georgiev, O., Bauer, K., Stone, A., Hedges, R., Schaffner, W., Utermann, G., Sykes, B. & Pääbo, S. (1994) Molecular Genetic Analyses of the Tyrolean Ice Man. *Science* **264**, 1775–1778.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. & Pääbo, S. (2001) Ancient DNA. *Nat Rev Genet* **2**, 353–359.

Izagirre, N. & de la Rúa, C. (1999) An mtDNA analysis in Ancient Basque Populations: Implications for Haplogroup V as a Marker for a Major Paleolithic Expansion from Southwestern Europe. *Am J Hum Genet* **65**, 199–207.

Izagirre, N., Artiach, P., & de la Rúa, C. (2000) Morphological and Molecular Evolution in Prehistoric Skeletal Populations. In: *Archaeogenetics: DNA and the population prehistory of Europe*, Eds.Renfrew, C. & Boyle, K. (McDonald Institute for Archaeological Research, Cambridge, UK), pp 197–201.

Izagirre, N., Alzualde, A., Alonso, S., Paz, L., Alonso, A. & de la Rúa, C. (2005) Rare haplotypes in mtDNA: applications in the analysis of biosocial aspects of past human populations. *Hum. Biol.* (under revision)

Kruskal, J. B. (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**, 28–42.

Larruga, J. M., Díez, F., Pinto, F. M., Flores, C. & González, A. M. (2001) Mitochondrial DNA characterisation of European isolates: The Maragatos from Spain. *Eur J Hum Genet* **9**, 708–716.

Maca-Meyer, N., Sánchez-Velasco, P., Flores, C., Larruga, J. M., González, A. N., Oterino, A. & Leyva-Cobián, F. (2003) Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet* **67**, 329–339.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B. & Torroni, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.

Mellards, P. A. (1992) Archaeology and the population-dispersal hypothesis of modern human origins in Europe. *Philos Trans R Soc Lond B Biol Sci* **337**, 225–234.

Menozzi, P., Piazza, A. & Cavalli-Sforza, L. L. (1978) Synthetic Maps of Human Gene Frequencies in Europeans. *Science* **201**, 786–792.

Mishmar, D., Ruiz-Pesini, E.,Golik, P.,Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., Sukernik, R. I., Olckers, A. & Wallace, D. C. (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* **100**, 171–176.

Montiel, R., Malgosa, A. & Francalacci, P. (2001) Authenticating ancient human mitochondrial DNA. *Hum Biol* **73(5)**, 689–713.

Otte, M. (1990) in *The World at 18,000 BP*, eds.Soffer, O. & Gamble, C. (Unwin, London) **Vol I**: 54–68.

Plaza, S., Calafell, F., Helal, A., Bouzerna, N., Lefranc, G., Bertranpetit, J. & Comas, D. (2003) Joining the Pillars of Hercules: mtDNA Sequences Show Multidirectional Gene Flow in the Western Mediterranean. *Ann Hum Genet* **67**, 312–328.

Pult, I., Sajantila, A., Simainen, J., Georgiev, O., Schaffner, W. & Pääbo, S. (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biol Chem Hoppe Seyler* **375**, 837–840.

Reynolds, J. B., Weir, B. S. & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**, 767–779.

Richards, M. B., Corte-Real, H., Foster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Balndelt, H. J. & Sykes, B. (1996) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* **59**, 185–203.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Gölge, M., Dimitrov, D., Hill, E., Bradley,

D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zahery, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A. & Bandelt, H. J. (2000) Tracing European Founder Lineages in the Near Eastern mtDNA Pool. *Am J Hum Genet* **67**, 1251–1276.

Richards, M., Macaulay, V., Torroni, A. & Bandelt, H. J. (2002) In Search of Geographical Patterns in European Mitochondrial DNA. *Am J Hum Genet* **71**, 1168–1174.

Richards, M. (2003) The Neolithic Invasion of Europe. *Ann Rev Anthropol* **32**, 135–162.

Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Avalantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckamn, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Corte-Real, H. B., de Knijff, P., Decorte, R., Dubrova, Y. E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T. A., Mikelsaar, A. V., Mitchell, R. J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C. & Jobling, M. A. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67**, 1526–1543.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V. & Wallace, D. C. (2004) Effects of Purifying and Adaptative Selection on Regional Variation in Human mtDNA. *Science* **303**, 223–226.

Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* **20**, 278–280.

Semino, O., Passarino, G., Oefner, P. J., Lin, A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A., Cavalli-Sforza, L. L. & Underhill, P. A. (2000) The genetic legacy of paleolithic homo sapiens sapiens in extant europeans: A Y-chromosome perspective. *Science* **290**: 1155–1159.

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., Zhivotovsky, L. A., King, R., Torroni, A., Cavalli-Sforza, L. L., Underhill, P. A. & Santachiara-Benerecetti, A. S. (2004) Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. *Am J Hum Genet* **74**, 1023–1034.

Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. & Barbujani, G. (2000) Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet* **66**, 262–278.

Sokal, R. R., Oden, N. L. & Wilson, C. (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* **351**, 143–145.

Straus, L. G. (1989) Age of Modern Europeans. *Nature* **342**, 476–477.

Sykes, B., Corte-Real, H. & Richards, M. (1996) Paleolithic and Neolithic contributions to the European gene pool. In: *Molecular Biology and Human Diversity* Eds.Boyce, A. J. Mascie-Taylor, CGN (Cambridge University Press, Cambridge, England), pp. 130–140.

Torroni, A., Bandelt, H. J. D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonné-Tamir, B. & Scozzari, R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* **62**, 1137–1152.

Vernesi, C., Caramelli, D., Dupanloup, I., Bertorelle, G., Lari, M., Capellini, E., Moggi-Cecchi, J., Chiarelli, B., Castrì, L., Casoli, A., Mallegni, F., Lalueza-Fox, C. & Barbujani, G. (2004) The Etruscans: A Population-Genetic Study. *Am J Hum Genet* **74**, 694–704.

Wallace, D. C., Ruiz-Pesini, E. & Mishmar, D. (2003) MtDNA variation, Climatic Adaptation, Degenerative Diseases, And Longevity. *Cold Spring Harb Sym* **63**, 471–478.

Wilson, J. F., Weiss, D. A., Richards, M., Thomas, M. G., Bradman, N. & Goldstein, D. B. (2001) Genetic evidence for different male and female roles during cultural transition in the British Isles. *Proc Natl Acad Sci USA* **98**, 5078–5083.

Xiao, F. X., Yotova, V., Zietkiewicz, E., Lovell, A., Gehl, D., Bourgeois, S., Moreau, C., Spanaki, C., Plaitakis, A., Moisan, J. P. & Labuda, D. (2004) Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts. *Eur J Hum Genet* **12**, 301–311.

Zvelebil, M. (1989) On the transition to farming in Europe, or what was spreading with the Neolithic: a reply to Ammerman. *Antiquity* **63**, 379–383.

Zvelebil, M. (1998) Genetic and cultural diversity of Europe: a comment on Cavalli-Sforza. *J Anthropol Res* **54**, 411–417.

Received: 26 June 2004
Accepted: 31 January 2005

SUPPLEMENTARY MATERIAL

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/ahg/ahg170sm.htm

This contains detailed information on each individual that was genetically analysed in the historical population of Aldaieta.