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 inthe Basque Country display clear differentiation in relation to all oth 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

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

M

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

 © University College London 2005

F **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).

M

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.

C , S P D S

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.

E DNA U P /G

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

 © University College London 2005

T. 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.

A RFLP C R M DNA

In order to classify the variability of the mitochondrial genome of the individuals buried in Aldaieta 61B10 1epnI03meedb365.5(or)TDNA 8994 H-aggtggcctgcagta08991 Tf 10.9589 0 0 10.9589 65.4247 340.7158 Tm [(Aa)-3015(A)(e)0(e(8991 Tfm6gr)TDN

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

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.

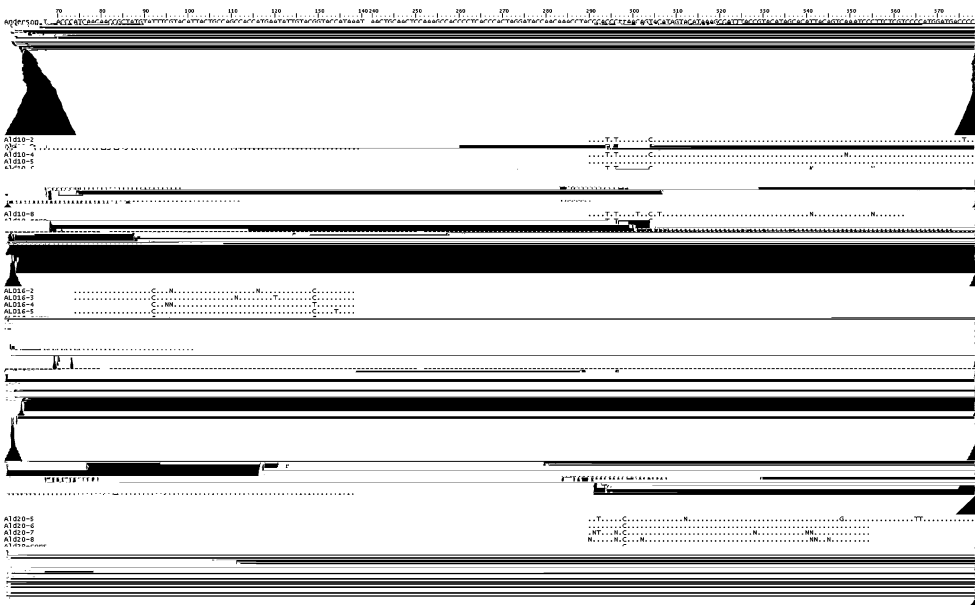
T 5 Number of individuals from Aldaieta following the correction of frequencies for kinship determined by RFLPs and by sequencing

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.

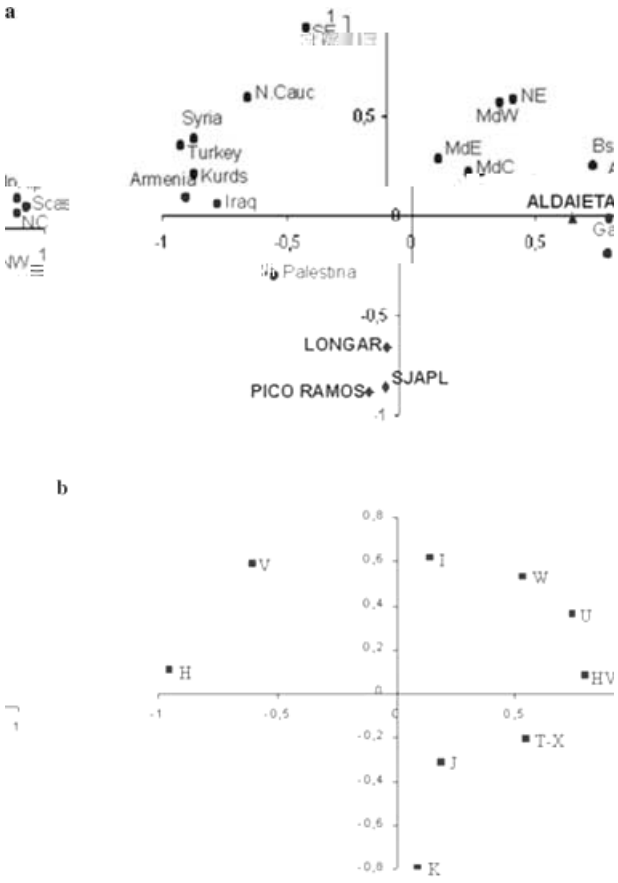
P , • C , , (PC) A

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 &



F 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).

© University College London 2005


a
N.Cauc
NE
MdW
Syria
Turkey
MdE
MdC
Armenia
Kurds
Iraq
ALDAIETA
Palestina
LONGAR
SJAPL
PICO RAMOS
b
V
I
W
U
H
HV
T-X
J
K

F t 4

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

Te

populations reflect the evolutionary processes experi-

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.

Copyright of Annals of Human Genetics is the property of Blackwell Publishing Limited. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.