

The Genetics of the Pre-Roman Iberian Peninsula: A mtDNA Study of Ancient Iberians

M. L. Sampietro¹, D. Caramelli², O. Lao¹, F. Calafell¹, D. Comas¹, M. Lari², B. Agustí³, J. Bertranpetit¹ and C. Lalueza-Fox^{1,*}

¹Unitat de Biologia Evolutiva, Departament Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona

²Dipartimento di Biologia Animale e Genetica, Laboratori di Antropologia, Università di Firenze, Italy

³Museu d'Arqueologia de Catalunya, Girona

Summary

The Iberians developed a surprisingly sophisticated culture in the Mediterranean coast of the Iberian Peninsula from the 6th century BC until their conquest by the Romans in the 2nd century BC. They spoke and wrote a non-Indo-European language that still cannot be understood; their origins and relationships with other non-Indo-European peoples, like the Etruscans, are unclear, since their funerary practices were based on the cremation of bodies, and therefore anthropology has been unable to approach the study of this people. We have retrieved mitochondrial DNA (mtDNA) from a few of the scarce skeletal remains that have been preserved, some of them belonging to ritualistically executed individuals. The most stringent authentication criteria proposed for ancient DNA, such as independent replication, amino-acid analysis, quantitation of template molecules, multiple extractions and cloning of PCR products, have been followed to obtain reliable sequences from the mtDNA hypervariable region 1 (HVR1), as well as some haplogroup diagnostic SNPs. Phylogeographic analyses show that the haplogroup composition of the ancient Iberians was very similar to that found in modern Iberian Peninsula populations, suggesting a long-term genetic continuity since pre-Roman times. Nonetheless, there is less genetic diversity in the ancient Iberians than is found among modern populations, a fact that could reflect the small population size at the origin of the population sampled, and the heterogenic tribal structure of the Iberian society. Moreover, the Iberians were not especially closely related to the Etruscans, which points to considerable genetic heterogeneity in Pre-Roman Western Europe.

Keywords : ancient DNA, mitochondrial DNA, Iberian Peninsula

Introduction

The so-called Iberian culture is considered to be the first historic culture of the Iberian Peninsula (to which it gave its name); it developed in the South and East along the Mediterranean coast from the indigenous Bronze Age culture, from the 8th to 6th century BC onwards (Fig 1). Although considered to be a general unit, in fact Iberians

were a conglomerate of independent groups and tribes that had their own governance rules, and thus do not represent a single political entity; in fact, there is much evidence of conflict among the different Iberian groups (Ruiz & Molinos, 1998). The cultural influences of people from the East Mediterranean (Greeks and Phoenicians) shaped the distinctive artistic expression of these people (Harrison, 1988; Ruiz & Molinos, 1998). The Iberians had an urban society in which the cities, dominated by the military elite, were usually independent states relying in local agricultural resources and metallurgy. The arrival of the Romans from the 3rd century BC, and conflicts with the Carthaginians, represented a cultural and political disruption and the displacement

Present address: Unitat d'Antropologia, Dept. Biologia Animal, Universitat de Barcelona, Barcelona

*Corresponding author: Carles Lalueza-Fox, Unitat d'Antropologia, Dept. Biologia Animal, Universitat de Barcelona, Avda Diagonal 645, 08028 Barcelona (Spain) Telf. 34934021461. E-mail: clalueza@ub.edu

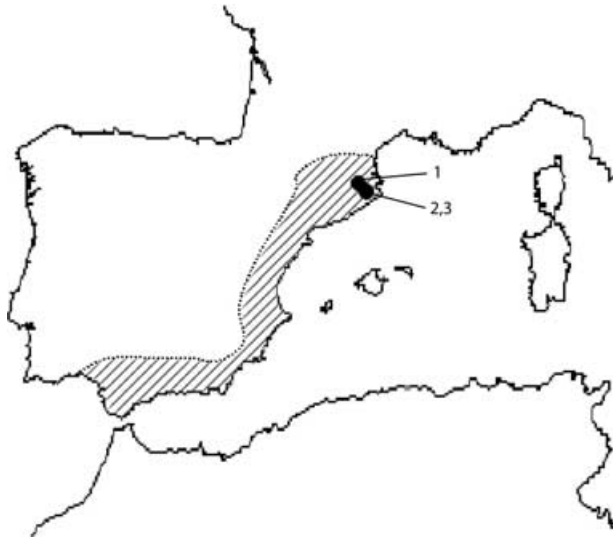


Figure 1 Area of expansion of the Iberian culture. Dots represent the archaeological sites where the samples studied were found (1: Mas Castellar, 2–3: Puig de Sant Andreu and Illa d'en Reixac).

of the centres of production to the coastal Roman settlements, resulting in an abandonment of the Iberian cities (Harrison, 1988).

The Iberians spoke a non Indo-European language that continued in use until Roman times, before being replaced by Latin. Numerous inscriptions made in this language, using a unique script, have been found; however, although the phonemes of the alphabetic characters can be transliterated, the actual language still cannot be understood. It was formerly thought that Basque, also a non Indo-European language, was closely related to Iberian, but now most scholars believe that both languages are separate, and probably represent part of a heterogeneous pre-Roman linguistic substratum (de Hoz, 1995). The Iberian language has also been related to Etruscan, another a non-Indo-European language dating to the end of the 8th century BC in Central Italy (Etruria, current Tuscany), but there is no clear evidence of this relationship (de Hoz, 1995).

It has been hypothesized that Iberians and Basques (and maybe Etruscans) should share genetic affinities as representatives of pre-Roman populations in Western Europe. The fact that these three populations spoke non Indo-European languages may also indicate the persistence of a Paleolithic substratum, according to the model of a demic expansion associated with agriculture and Indo-European languages from the Middle East. The

distinctiveness of the Basques in the context of present-day Iberian populations has been detected with classical genetic markers, and has been interpreted as the result of long-standing isolation (Bertranpetit & Cavalli-Sforza, 1991); nonetheless, a general genetic replacement of pre-Indo-European populations during the Neolithic remains controversial, and other later events related to cultural changes have only been superficially explored. Therefore, genetic analyses of these groups are of great interest to help to understand the phylogeographic structure of past and modern Western European populations. Unfortunately, the Iberians incinerated their dead, thus making their numerous necropolises unsuitable for ancient DNA analysis. However, some skeletal remains are occasionally found, especially in large towns such as Ullastret (North East of Catalonia); some of them are related to a ritual practise of nailing skulls, in which the heads of the enemies were displayed in public places with a long nail going all the way through the skull from the forehead to the cranial base. In this study, we have analyzed some of these nailed skulls plus other Iberian remains to obtain, for the first time, a genetic picture of the human pre-Roman substratum in the north-east of the Iberian Peninsula.

Genetic studies have shown that mitochondrial DNA (mtDNA) substitutions that have accumulated along maternal lineages or haplogroups have diverged as human populations expanded through different continents; these lineages can be defined by particular SNPs in the coding region of the mtDNA, correlated to different haplotype sequences in the hypervariable control region (HVR-1) of the mitochondrial genome. Molecular analyses have revealed the existence of Eurasian and European specific haplogroups (named H, I, J, K, T, U, V, X and W); their distribution and characterization can help reconstruct past human migrations and population affinities in Europe, as well as provide a genetic context for ancient DNA studies. We have obtained data from both HVR-1 haplotype sequences and haplogroup-diagnostic SNPs from Iberian individuals by means of ancient DNA techniques.

Materials and Methods

Twenty tooth samples belonging to seventeen different individuals were analyzed; incomplete or fragmented

Table 1 Samples and sites analyzed

Sample	Site	Sample code	Sex
1	Puig de Sant Andreu (Ullastret)	UE 14015	?
2	Puig de Sant Andreu (Ullastret)	UE 14045	?
3	Illa d'en Reixac (Ullastret)	UE 15112	M
4	Mas Castellar (Pontós)	1000012-2	?
5	Illa d'en Reixac (Ullastret)	UE 1070	M
6	Puig de Sant Andreu (Ullastret)	N.I. 3967	?
7	Illa d'en Reixac (Ullastret)	15091	?
8	Puig de Sant Andreu (Ullastret)	UE 14020	?
9	Puig de Sant Andreu (Ullastret)	T.C. 4 E I-II	?
10	Puig de Sant Andreu (Ullastret)	UE 14015-1	M?
11	Illa d'en Reixac (Ullastret)	UE 15224	M
12	Puig de Sant Andreu (Ullastret)	UE 14119	?
13**	Illa d'en Reixac (Ullastret)	UE 15078-1	M
14	Puig de Sant Andreu (Ullastret)	2002 Z13	?
15*	Puig de Sant Andreu (Ullastret)	3613	M
16*	Puig de Sant Andreu (Ullastret)	3615	M
17*	Puig de Sant Andreu (Ullastret)	1401	M

*: nailed skulls, **: signs of violence (possible decapitation).

teeth were discarded for the study. The remains were recovered from three different archaeological sites in Girona (Catalonia, North Eastern Spain): Puig de Sant Andreu ($n = 11$), Illa d'en Reixac ($n = 5$) and Mas Castellar ($n = 1$). Mas Castellar is a small Iberian village, around 15 km inland, which was inhabited from the 7th to the 2nd century BC (Adroher *et al.* 1993). Illa d'en Reixac is a settlement established around 600 BC on the shore of an ancient, now dessicated, lagoon, later abandoned for a neighbouring settlement, Puig de Sant Andreu, which is part of the Ullastret archaeological site, one of the largest Iberian towns (Martín, 1985; Harrison, 1988). Ullastret flourished between 535 and 200 BC, partly due to the commercial influences of the Greek coastal city of Emporium, in an area dominated by the *Indiketes* Iberian tribe. Ullastret lies on a hill; the city comprised 11 square kilometres, was surrounded by solid walls at least four metres high, regularly reinforced with rounded, massive towers, and had an acropolis on its top (Martín, 1985). The city was destroyed by the Romans around 200 BC and abandoned.

With the exception of three fairly complete skulls, most of skeletal samples consist of fragmentary maxillary and mandibular remains that were not found in a funerary context (Table 1). At least three of the skulls correspond to ritualistically executed individuals (nailed skulls) from Ullastret, while another one is a mandible

that shows signs of violence, probably associated with decapitation (Agustí, 1999). All the remains that have been sexed using osteological traits have been considered adult males.

Extraction and Amplification

The root tip of each tooth was sectioned and the crown glued back to its alveolus; the discarded root was used for DNA extraction. In two cases, it was possible to duplicate the sample for independent replication undertaken at the University of Florence. Tooth root surfaces were first cleaned with bleach and then ground to powder; the samples weighed from 0.1g to 0.5g. The extraction method has been described elsewhere (Caramelli *et al.* 2003). In brief, ten millilitres of EDTA (pH: 8; 0.5M) were added to the powder overnight to remove mineral salts; after centrifugation, the EDTA was carefully poured off and the remaining sample was incubated overnight at 50°C in a lysis solution (1ml SDS 5%, 0.5 ml TRIS 1M, 8.5 ml H₂O and proteinase K). Then the sample was extracted with phenol, phenol-chloroform and chloroform-isoamyl alcohol and concentrated with centricons (Millipore) up to a 50–100 µl volume.

Extraction procedures were performed in an isolated pre-PCR area exclusively dedicated to ancient DNA studies, physically isolated from the main laboratory, with positive air pressure, overnight UV light and frequent bench cleaning. All samples and reagent manipulations were performed in a laminar flow cabinet routinely irradiated with UV light. Negative controls were used throughout the extraction and amplification analysis. To help avoid laboratory contamination, sterile, aliquoted reagents, sterile gloves, sterile pipettes, facemasks, cover-all coats and filter pipette tips were used.

The mtDNA HVR1 region was amplified in different overlapping fragments with sizes ranging from 232 to 94pb depending on the state of preservation of the samples. For this analysis several primer pairs were used (Table 2). Other additional primer pairs were designed to amplify the region where the most common European haplogroup specific sites are present (Table 2); this allowed us to ascertain by sequencing or enzymatic digestion some diagnostic SNPs that assigned a

Table 2 Primers used in the study

HVR1 primers
L16022 5'CTAATTTAAACTATTCTCT 3'
L16055 5'GAAGCAGATTTGGGTACCAC 3'
L16081 5'TTGACTCACCCATCAACAA 3'
L16122 5'CATTACTGCCAGCCACCATGAATA3'
L16131 5'CACCATGAATATTGTACGGT3'
L16185 5'AACCCAATCCACATCAAAAACC3'
L16209 5'CCCATGCTTACAAGCAAGT 3'
L16223 5'AGCAAGTACAGCAATCAAC 3'
L16261 5'CAACTATCACACATCAACTGCAA3'
L16347 5'CGTACATAGCACATTACAGT 3'
H16142 5'ATGACTACAGGTGGTCAAG 3'
H16158 5'TGTGGATTGGGTTTTTATGT 3'
H16211 5'TAGTTGAGGGTTGATTGCTG 3'
H16218 5'TGTGTGATAGTTGAGGGTTG 3'
H16247 5'CAACTATCACACATCAACTGCAA3'
H16281 5'TTAAGGTTGGGTAGGTTTGT 3'
H16356 5'GTCATCCATGGGGACGAGAA 3'
H16385 5'GGTGGTCAAGGGACCCCTAT 3'
H16401 5'TGATTTACGGAGGATGGTG 3'
Coding region SNPs
L12227 5'GAAAGCTCACAAGAACTGC3'U (12308)
H12341 5'GGTTATAGTAGTGTGCATGG3'
L 6999 5'CAAACCTACTACTAGACATCG3'H (7028)
H 7066 5'GAATGAAGCCTCCTATGATGG3'
L13669 5'CACCCTTACTAACATTAACG3'J (13708)
H13725 5'TAGTAATGAGAAATCCTGCG3'
L13257 5'AATCGTAGCCTTCTCCACTTCA3'T (13368)
H13372 5'TTGTTAAGTTGTGGATGATG3'
L2993 5'TAGGGTTTACGCCTCGATGT3'
H3054 5'TCTGAACTCAGATCACGTAGGACT3'
H1 (3010A)

Some HVR1 primers have been published elsewhere (Handt *et al.* 1996; Stone & Stoneking, 1998; Caramelli *et al.* 2003); others have been designed for this study. Coding region primers were designed in the present study.

sample unequivocally to a haplogroup in the mtDNA gene genealogy. PCR amplifications were performed in 25 μ l reactions with 1 μ l to 3 μ l of extract (depending on the extract volume and/or the presence of inhibitors), 1.2 U of *Taq* DNA polymerase (Ecogen), 1X reaction buffer (Ecogen), 1.4 mg/ml BSA, 2.5–2.1 mM MgCl₂, 0.2 mM dNTP's and 1 μ M of each primer. The PCR reactions were subjected to 40 amplification cycles (1 min step at 94°C, 1 min step at 50°C and 1 min step at 72°C) with an initial denaturing step at 94°C for 5 min and a last elongation step at 72°C for 7 min. Products were electrophoresed in 1.6% low-melting point agarose gels (Invitrogen) stained with ethidium bromide. Bands visu-

alized under UV light were excised from the gel, melted at 65°C, diluted in 100–150 μ l of double-distilled water and heated for one hour at 65°C. Those products became the template of a further 35 cycles of PCR, with limiting reagents; the bands were subsequently purified with GFX columns (Amersham Biosciences) and sequenced with an ABI 3100® DNA sequencer (Applied Biosystems), following the supplier's instructions.

Uracil-N-Glycosylase (UNG) Treatment

10 μ l of DNA extract for some of the samples (Table 3) was treated with 1U of Uracil-N-Glycosylase(UNG) for 30 min at 37°C to excise uracil bases caused by the hydrolytic deamination of cytosines. UNG reduces sequence artefacts caused by this common form of post-mortem damage, resulting in apparent G/C→A/T mutations and subsequent errors in the sequence results (Hofreiter *et al.* 2001). After this treatment, extracts were subjected to the same PCR conditions as described above.

Cloning of PCR Products

Direct sequences with heteroplasmic positions were either sequenced and/or cloned (Table 3). If an observed base change is due to a random DNA damage, the likelihood of finding this change after multiple (two or three) amplifications can be estimated to be less than 0.1% (Hofreiter *et al.* 2001). Therefore, most of the samples were amplified and sequenced several times, using different sets of primers (Table 3); fragments with discrepant results were cloned. PCR products were treated with a pMOSBlue blunt ended cloning kit (Amersham Biosciences) following the manufacturer's instructions. Seven microlitres of PCR product were treated with pK enzyme mix, incubated at 22°C for 40 min and ligated into the pMOSBlue vector overnight. Two μ l of the ligation product were transformed into 40 μ l of competent cells, grown in 160 μ l of SOC medium at 37°C for one hour and plated on IPTG/X-gal agar plates. After 16 hours, white colonies were subjected to direct PCR screening using the T7 and U-19 universal primers. Inserts that yielded the correct size were identified by agarose gel electrophoresis, purified and sequenced.

Table 3 Amplification strategy of the analyzed samples

Sample	Additional typing	HVR1 fragments sequenced
1	C, U	055-142/055-218/122-356/209-356/209-401
2	C, U, E	055-142/055-218/209-356/209-401
3	A, C, I, U	055-142/081-281/122-218/131-211/209-401/223-385
4	Q, U	055-142/122-218/131-211/209-401
5	A, I, Q	022-158/055-142/081-281/131-211/209-356/209-401
6	Q	055-142/131-211/209-356
7		131-211/209-356/209-401
8		055-142/131-211/209-356/209-401
9	C	055-142/131-211/209-356/209-401
10		055-142/131-211/209-401
11	C, Q	055-142/131-218/209-356/209-401/347-401
12	Q	055-142/131-218/209-356/209-401/347-401
13	C	055-142/209-401/347-401
14		055-142/131-211/209-401/247-356/347-401
15	C	055-142/131-211/185-356
16	C	055-142/055-211/131-211
17		055-142/131-211/185-261/209-401/347-401

I, Sample independently replicated; C, cloned; E, multiple extractions from the same individual; A, aminoacid analysis; U: samples with UNG treatment; Q: quantitation DNA analysis.

Quantitation of Number of DNA Template Molecules

To find out if the amount of mtDNA template was large enough to allow us to obtain reproducible results, a Real Time-PCR experiment was performed. We obtained a DNA standard by PCR-amplification of the HVR1 region (Anderson *et al.* 1981) from an anonymous laboratory donor with known sequence; the PCR product was purified and quantified by a Nanodrop Spectrophotometer instrument. Two different sized fragments (107bp and 278bp) within the HVR1 mtDNA region were assayed to estimate the mtDNA preservation. The design of primers and probes was performed with Primer Express 2.0 software (Applied Biosystem). The primer sequences for the small fragment were: L16001 ACCATTAGCACCCAAAGCTAAGA and H16065 GCGGTTGTTGATGGGTGAGT, and for the larger one: L16088 TCACCCATCAACAACCGCTAT and H16344 GGGACGAGAAGGGATTTGACT. The probe oligonucleotide sequence for the small fragment was FAM-CAAGCAAGTACAGCAA-MGB and for the large was VIC-GAAGCAGATTTGGGTAC-MGB (Alonso *et al.* 2004). Real-time PCR amplification was performed in a 20 μ l reaction with 1x reaction TaqMan Universal PCR Master Mix (Applied Biosystem), 0,5

μ M each primer, 50nM probes, 1mg/ml BSA and 1 μ l DNA extract. Ten-fold serial dilutions of the purified and quantified standard were included in the experiment to create a standard curve, in order to quantify the number of initial mitochondrial DNA molecules of each size in the Iberian samples.

Population Analysis

Combining the information obtained from the HVR1 sequence and the different diagnostic coding region SNPs, each sample was classified into its corresponding mitochondrial DNA haplogroup (Table 4); the haplogroup frequencies were crucial for exploring the possible phylogeographic structure of the Iberian population (Table 5). To explore comparatively the haplotype composition, a correspondence analysis was generated with STATISTICA software (StatSoft, Inc. 2001 version 6); because it has been suggested that Punic and Eastern influences could have contributed to the genesis of the Iberian culture, we have included in the analysis three populations from North Africa (including one from the present-day area of Carthage) plus two more from Central Italy (including one from the present-day region that was ancient Etruria). A pairwise distance matrix was calculated with these populations

HAPLOGROUP	SAMPLE	HVR1 HAPLOTYPE	SNPs
H	1	16287	n.d.
H	5, 15	16126 16311	- 7025 AluI (#5), 7028 C (#15) 3010 G, 13708 G
H	6, 11, 17	16126	7028 C, 3010 G, 13708 G
H1	7	CRS	- 7025 AluI, 3010 A
H1	9	16126	- 7025 AluI, 3010 A
H	12	16273	- 7025 AluI
Pre-HV	10	16362	7028 T, 14766 C
J*	4	16126 16189	- 13704 BstNI, 3010 G
J*	16	16126	13708 A, 7028 T
K*	13	16224 16311	n.d.
T	8	16126 16294	n.d.
U4	2	16356	7028 T, 12308 G, 3010 G
U5*	14	16270 16281	n.d.
U5a	3	16256 16270	n.d.

Table 4 Haplogroup attribution and HVR1 haplotypes in ancient Iberians

n.d.: No determined.

Table 5 Haplogroup frequencies (%) in Western Mediterranean Populations.

Pop (n)	Alg (47)	MB (64)	Tun (47)	And (158)	Bas (173)	Cat (78)	CS (50)	Gal (103)	Val (30)	NPo (100)	CPo (82)	SPo (59)	Ibe (17)	CIIt (83)	Tus (49)
L1	6.4	1.6	-	-	-	-	2.0	-	-	1.0	1.3	1.7	-	-	-
L2	6.4	-	12.8	0.6	-	-	2.0	1.0	3.4	3.0	2.4	1.7	-	-	-
L3	14.9	1.6	14.9	1.3	-	1.3	-	1.0	-	1.0	2.4	1.7	-	1.2	2.0
D	-	-	-	-	-	-	-	-	-	-	-	-	-	1.2	-
M1	12.8	-	4.2	1.9	-	-	-	-	-	-	1.3	-	-	-	-
M5	-	-	-	1.9	-	-	2.0	-	-	-	-	-	-	-	-
N	-	-	-	0.6	-	1.3	-	-	6.6	-	1.3	-	-	-	-
I	-	-	2.1	1.9	-	-	-	-	-	1.0	-	1.7	-	-	4.1
W	-	-	-	1.3	-	5.1	2.0	1.9	-	2.0	1.3	-	-	1.2	2.0
X	2.1	-	2.1	3.2	1.7	2.6	2.0	1.0	-	-	3.6	1.7	-	3.6	6.1
R1	-	-	-	-	1.7	-	-	-	-	-	-	-	-	-	-
J/T	-	1.6	-	0.6	-	-	-	-	-	-	-	-	-	-	-
T	4.2	15.6	6.4	4.4	5.2	7.7	10.0	2.9	-	11.0	11.0	10.2	5.9	15.7	10.2
J	12.8	9.4	4.2	7.0	4.0	2.6	8.0	9.7	6.6	6.0	6.1	8.5	11.8	6.0	14.3
U*	2.1	6.2	6.4	10.1	13.2	9.0	20.0	12.6	16.6	16.0	14.6	11.9	17.6	9.6	12.2
U6	-	7.8	4.2	1.9	-	-	2.0	1.9	-	7.0	-	-	-	-	-
K	4.2	7.8	6.4	6.3	5.2	6.4	2.0	3.9	10.0	3.0	7.3	6.8	5.9	7.2	8.2
HV	-	-	6.4	2.5	-	1.3	2.0	1.0	3.4	-	-	-	5.9	1.2	2.0
H	34.0	42.2	23.4	46.2	57.8	56.4	46.0	59.2	53.3	41.0	37.8	44.1	52.9	47.0	38.8
V	-	6.2	-	5.7	10.4	5.1	-	2.9	-	8.0	7.3	6.8	-	4.8	-
Other	-	-	6.4	2.5	0.6	1.3	-	1.0	-	-	2.4	3.4	-	1.2	-

Alg: Algerians; MB: Moroccan Berbers; Tun: Tunisians; And: Andalucians; Bas: Basques; Cat: Catalans; CS: Central Spanish; Gal: Galicians; Val: Valencians; NPo: North Portuguese; CPo: Central Portuguese; SPo: South Portuguese; **Ibe: Ancient Iberians**; CIIt: Central Italians; Tus: Tuscans; (*): Excluding U6. N includes sequences carrying the HVR1 substitutions diagnostic of either N1a or N1b.

and was visualised by means of MDS (MultiDimensional Scaling). This method allows us to represent a distance matrix in a number of dimensions specified *a priori*. The degree of agreement between the original pairwise distance matrix and the matrix of the MDS representation is evaluated by the stress measure; as in the corre-

spondence analysis, the populations are represented in a two-dimensional space.

The samples considered for comparison were from the Iberian Peninsula, plus some from North Africa and Italy which, according to some scholars, are regions that could have contributed to the genetic background

of the ancient Iberians: Algerians, Tunisians and Moroccan Berbers (Plaza *et al.* 2001; Rando *et al.* 1998); Andalusians, Basques, Catalans, Central Spanish, Galicians and Valencians (Bertranpetit *et al.* 1995; Corte-Real *et al.* 1996; Crespillo *et al.* 2000; López-Soto *et al.* 2000; A. Alonso (personal communication); Plaza *et al.* 2001; Richards *et al.* 2000; Salas *et al.* 1998); North, Central and South Portuguese (Pereira *et al.* 2000); Central Italians and Tuscans (Tagliabracci *et al.* 2001; Francalacci *et al.* 1996). Also, several parameters related to population internal structure (nucleotide diversity, sequence diversity and mean pairwise differences) were estimated for the Iberians using the Arlequin 2000 software (Schneider *et al.* 1996). In addition, the Iberian sequences were placed in a general network of Eurasian HVR-1 sequences, constructed using a reduced-median algorithm (Bandelt *et al.* 1995), implemented in the Network 3.0 program.

Results

Authentication Criteria

The main authentication criteria proposed as quality standards in ancient DNA research (Cooper & Poinar, 2000; Poinar, 2004) were followed in order to obtain empirical evidence about how favourable the Ullastret site environment was for DNA preservation 1). The analysis was carried out in isolated, dedicated ancient DNA laboratories, both in Barcelona (Unitat de Biologia Evolutiva, Universitat Pompeu Fabra) and Florence. 2) The degree of aminoacid racemization was analysed in two representative samples, following the procedures described in Vernesi *et al.* (2001); the values (<0.10 for aspartic acid) obtained (Table 6) are suggestive of good biochemical preservation and therefore, compatible with DNA survival (Poinar *et al.* 1996; Poinar, 2004). Interestingly, the alanine D/L and glutamine D/L val-

ues are lower than the aspartic acid D/L; because the racemisation of Alanine and Glutamine is slower than that of aspartic acid, this difference is a criterion for distinguishing endogenous protein preservation from putative modern contaminants (Poinar *et al.* 1996). Although this test was only applied to a small subset of the samples (~12%), it must be taken into consideration that most of the samples come from a very specific area of the Ullastret site, and it is therefore likely that all samples have been subjected to similar taphonomic processes. 3) As mentioned above, some fragments were cloned, to distinguish sequence heterogeneities due to DNA damage and subsequent amplification errors, and also to reveal putative multiple sequences due to a mixture of endogenous sequences and contaminants. The *Taq* misincorporation error values obtained (number of substitutions per 1000 bp) are low and yielded values between 0 and 5.68, similar to those observed in other ancient DNA studies (Cooper *et al.* 2001; Caramelli *et al.* 2003). 4) Two samples (Table 3) were independently extracted, amplified and cloned in a different laboratory (Florence), as a control for possible intra laboratory contamination; in both cases, the substitutions observed were the same as those found in the Barcelona laboratory (Table 4, see also Table 1 of supplementary material). A C to T substitution at nucleotide 16,192 in sample 3, found in all the Florence clones, was not observed in two different amplifications in Barcelona, with different sets of primers, and therefore it has been quoted as putative postmortem damage due to cytosine deamination (Hofreiter *et al.* 2001) in the haplotype table (Table 4). Also, one sample (number 2) underwent two different extractions in different periods of the study; the sequences obtained from both extracts (with a T to C transition at position 16,356) were identical (Table 3). In addition, none of the sequences obtained were identical to the mtDNA haplotypes of the three researchers (M.L.S., C.L.-F, and D.C) directly involved in the analytical procedures. 5.) PCR and extraction blanks were included to detect any possible internal contamination, and also different sets of overlapping primers were used to be sure that the results were reproducible. 6) Quantitation of the copy number of mtDNA by RT-PCR showed that the number of DNA molecules/ μ l in the samples analyzed was high enough (Table 7) to ensure reproducibility of the results; in assays of

Table 6 Aminoacid racemisation analysis of two Iberian samples

	D/L Alanine	D/L Aspartic	D/L Glutamine
Sample 3	0.005	0.06	0.008
Sample 5	0.004	0.038	0.006

Values lower than 0.1 in the D/L Aspartic ratio are compatible with DNA preservation.

Table 7 Quantitation of DNA in the extracts by RT-PCR analysis.

Sample	Ct	Molecule's number of 107bp fragment (95%)
4	32.72	371–870
5	32.43	463–1088
6	32.28	520–1221
11	31.94	676–1586
12	32.89	325–763

The cycle threshold parameter (Ct) was determined by the SDS software as the fractional cycle number at which the fluorescence increases exponentially. The number of molecules (95% confidence interval) in each fragment is calculated by the standard curve (from 3.7×10^6 to 376 copies; from 20.13 to 31.98 Ct). No PCR products of 327 bp were obtained. One to three μ l of extract were used for each PCR.

competitive PCR, a figure of >1000 template DNA molecules has been proposed as a value that could be expected to yield reproducible results (Handt *et al.* 1996), while <300 could produce irreproducible sequences (Poinar, 2004); however, RT-PCR offers a new approach to the quantitation of template molecules with more precise figures of potential reproducibility (Alonso *et al.* 2003). In addition, the samples showed appropriate molecular behaviour for ancient extracts because it was not possible to amplify a fragment longer than 327 bp.

Iberian Sequences

Thirteen different sequences were present in the 17 samples analyzed; they are listed in Table 5 according to their haplogroup assignment. It was possible to sequence the mtDNA HVR1 from positions 16,055 to 16,401 (according to Anderson *et al.* 1981) in almost all the samples, with only a few gaps in samples 7, 13 and 16 (Table 5). This could slightly alter the haplotypes found in those sequences, although the general population inferences would not be significantly affected.

Three samples showed the C to T position 16,126 substitution associated with a C at position 7,028 (and therefore, belonging to the H haplogroup), while another one had again the 16,126 substitution, but this time associated with a T at 7,028 and a A in 13,708 (and therefore, belonging instead to the J haplogroup). This example shows the importance of typing diagnostic haplogroup positions in mitochondrial analysis. In

addition, two samples from the H haplogroup showed a combination of 16,126 and C to T at 16,311, while two others had the CRS sequence. Most of the haplotypes have already been described in European populations, with the exception of the G to A transition at 16,273 of sample 12; however, it has been detected in two different amplifications (16,209–16,356 and 16,209–16,401), and thus, it is unlikely that it could be attributed to post-mortem DNA damage (Hofreiter *et al.* 2001). In addition, the haplotype from sample 4 (16,126 and 16,189) has always been described in association with a C to T transition at 16,069 in J haplogroup sequences; we couldn't detect this substitution in the 16,055–16,142 fragment. The 16,069 position is a hotspot prone to DNA damage (Gilbert *et al.* 2003), but is also a mutational hotspot (Wakeley, 1993), and therefore we have no compelling evidence for considering it an artifact due to DNA degradation. In addition, the fragments treated with the Uracil-N-Glycosylase (UNG) enzyme yielded the same sequences as those without this treatment. However, the lack of the 16,189 substitution in one of the two amplifications that included this position suggest a potential source of error in this sample (one researcher has the 16,189 substitution, but also 16,185, which was not observed in any of the clones); at present, we cannot exclude the possibility that the endogenous sequence might lack the 16,189 position.

Another problematic sample is number 1. We found the C to T 16,287 substitution in only 7 out of 12 clones; however, it is present in three different amplifications (16,122–16,356; 16,209–16,356; 16,209–16,401). In this situation, template damage is an unlikely explanation, as it would need to affect the same position independently three times. Some of the clones that show a C at 16,287 seem to carry a high number of additional substitutions (one with 16,134, one with 16,145, 16,176 and 16,222, and another one with 16,189), which could be interpreted as evidence of independent sources of low-level contamination. While this cannot be excluded, three of these additional substitutions (16,134, 16,145 and 16,176) are rare or have never been described in the modern databases; therefore, it is unlikely that they could represent contaminant haplotypes.

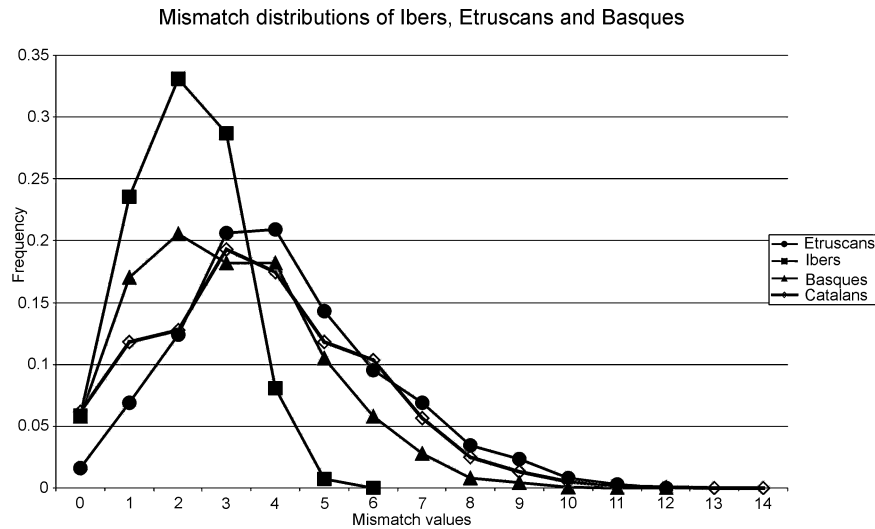


Figure 2 Mismatch distribution of Etruscans, Iberians, Basques and Catalans. Iberians show the least intrapopulation diversity.

The most frequent haplogroup is H (52.9%), followed by U (17.6%), J (11.8%), and pre-HV, K and T at the same frequency (5.9%). No samples were found to correspond to other haplogroups that are widely present in the Iberian peninsula populations (Table 7), such as V, X, I or W. The North African U6 subhaplogroup and Sub-Saharan African L lineages are also absent from the ancient Iberians analyzed so far; therefore, the possible entry of U6 lineages prior to the Muslim conquest in the 8th century A.D., as suggested by some authors, remains unproven. However, it is recognized that the sample size is at present too small to exclude any competing hypothesis about a possible North African genetic contribution to the genesis of the Iberian peninsula populations.

Phylogeographic Analysis

The Iberians shared several sequences with modern Western Mediterranean populations; the maximum ratio of shared haplotypes ($n = 5$) was with the Andalusians (the haplotypes are CRS; 16356; 16256–16270; 16362; 16224–16311); this is not surprising, since Andalusia was also an area where the Iberian culture developed. The ratio of sequences shared with other populations, both modern and ancient, is increasingly lower; however, our interest in shared sequences is small, not only because it depends strongly on the sample size but also because it is affected by the very low diversity within the Iberians.

One of the most striking findings was the low diversity found in the Iberian sample. It has the lowest mean pairwise difference of all European samples considered (mean equals 2.12) and the distribution is strongly shifted to the left, towards very small values (Fig 2). This does not seem to be due to a bias of small sample size; if we take 1,000 resamples of 17 sequences (the ancient Iberian sample size) from the Catalans and the Basques, only 1.2% of the former and 5.2% of the latter show mean values lower than that found for the ancient Iberians.

The correspondence analysis of the Western Mediterranean populations showed a marked geographic structure between North African and European populations (not shown); the Tunisians and the Algerians are separated from the rest due to the high frequency of L1, L2, L3 and M1 haplogroups, while the Moroccan Berbers and North Portuguese populations are separated because of the U6 and J/T haplogroups. In a second correspondence analysis (Fig 3), generated after removing the North African samples to explore the Western European populations in detail, the Basques and the North Portugal populations were the most divergent samples; the former are influenced by the high frequencies of R1 and V haplogroups, while the latter are influenced by high frequencies of L1, L2 and U6. There is not a clear structure among the Iberian peninsula populations; the ancient Iberians are clustered equidistant to present-day populations from Valencia

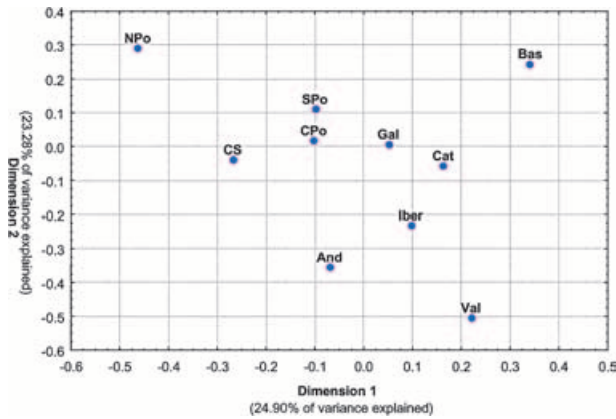


Figure 3 Correspondence analysis between populations from the Iberian peninsula; And: Andalucians; Bas: Basques; Cat: Catalans; CS: Central Spanish; Gal: Galicians; Val: Valencians; NPo: North Portuguese; CPo: Central Portuguese; SPo: South Portuguese; . Iber: Ancient Iberians.

and Catalonia (two of the present-day regions corresponding to the ancient Iberian culture) and Andalusia, in the centre of a triangle. A MDS analysis was also performed with the HVR-1 sequences; however, the high stress value (0.14) of the representation precluded taking it into consideration (not shown).

In the network analysis (Fig 4), it can be observed that the Iberian sequences, despite the small sample size, are distributed along a wide range of the main West Eurasian haplogroups, suggesting that the main lineages of present-day mitochondrial phylogenetic structure in Europe were already present in the Iberian peninsula in pre-Roman times; interestingly, many of the Iberian sequences are located in the central haplogroup and sub-haplogroup nodes (H1, T, J, U4 and K).

Discussion

Our statistical analyses have shown that ancient Iberians were not significantly different from modern populations from the same region (roughly, North-East and East of the Iberian peninsula) in haplogroup composition (Fisher’s exact test, $p = 0.9$). Despite biases due to the small sample size, the ancient Iberians have all the main European haplogroups present in modern Iberian peninsula populations; discrepancies between ancient Iberians and some populations such as those from Andalusia or Portugal are mainly due to the significant proportion of North and Sub-Saharan African lineages

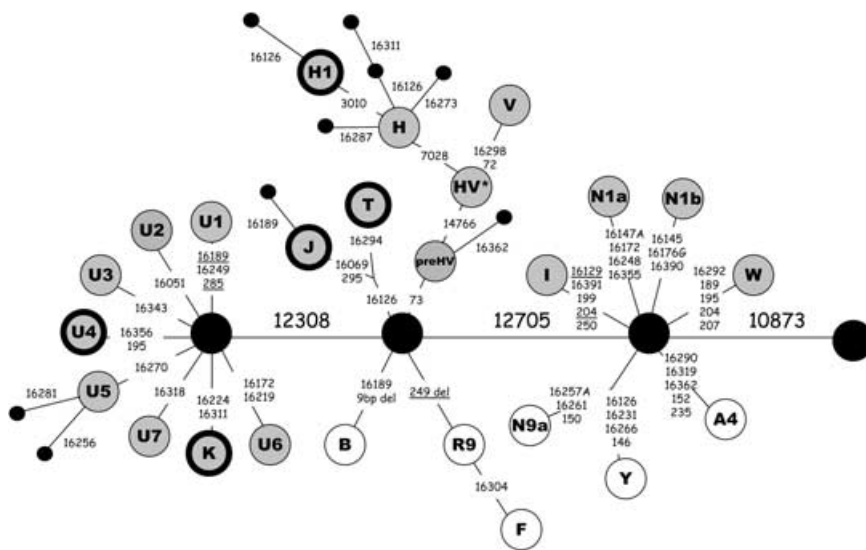


Figure 4 Phylogenetic reconstruction of the some of the major mitochondrial DNA haplogroups, as established by previous authors; the network has been simplified for clarity (L and M haplogroups are not included). Numbers along the links indicate substitutions, underlined numbers indicate recurrent mutations. West Eurasian, East Asian, Pan-Eurasian and African lineages are shown in grey, white, striped and dotted circles, respectively. The Iberian samples are displayed in dark circles (dark rings when are in the central nodes of some haplogroup).

in the latter populations. These lineages were probably introduced by relatively recent migrations during the Muslim period, from the VIIIth to XVth century A.D. (Bosch *et al.* 2001; Plaza *et al.* 2003). The lack of U6 lineages in the ancient Iberians, although not conclusive due to the small number of individuals analyzed, does not provide support for an earlier gene flow between North Africa and the Iberian Peninsula, as some authors have suggested, and supports the hypothesis of a modern, historical entrance of genes from North-Africa into Iberia.

Moreover, despite suggested similarities between Euskera (Basque language) and the enigmatic Iberian language (see, for instance, de Hoz, 1995), both groups do not seem to be especially close to one another from a genetic point of view. The Basques are a group with limited diversity and relatively high frequencies of the R1, H and V haplogroups (Richards *et al.* 1996, 2000; Torroni *et al.* 1998); although the frequency of haplogroup H in Iberians is >50%, no sequences belonging to R1 or V have yet been found among them. In addition, a high frequency of the H haplogroup is a trait shared by other modern Iberian populations, such as those from Catalonia, Valencia and Galicia (see Table 7).

The H haplogroup has recently been subdivided, and some of the sub-haplogroups (H1 and H3) show a geographic pattern compatible with post-glacial expansions from the Cantabrian-Basque region (Achilli *et al.* 2004); we have successfully typed the H1-defining SNP, a G to A transition at position 3010, in two samples (numbers 7 and 9). Future ancient DNA studies on European samples will be able to provide useful phylogeographic information with the additional typing of the different SNPs that subdivide the widespread H haplogroup.

Previous analysis of 121 teeth from four prehistoric sites from the Basque country (dated between ca 5,000 years BP to 3,400 years BP) determined the absence of the V haplogroup through RFLP typing of diagnostic SNPs in the ancient Basques (Izagirre & de la Rúa, 1999). Present-day Basque populations display frequencies of this haplogroup ranging from 3.3% to 20%, depending on the sample analyzed (Bertranpetit *et al.* 1995; Torroni *et al.* 1998; Richards *et al.* 2000); it has been suggested that this discrepancy may reflect the effect of drift associated with a Basque population substructure, or to

the posterior arrival of V haplogroup migrants (Izagirre & de la Rúa, 1999). In any case, the absence of V haplogroup sequences in the ancient Iberians suggests that this lineage was not especially prevalent in the ancient populations with non Indo-European languages from the Iberian Peninsula. It is likely that both Iberians and Basques were part of a complex mosaic of pre-Roman peoples occupying the Iberian Peninsula that probably emerged from the Bronze and Iron Age local cultures. Basques could represent an older, relic population that has subsisted up to present times with a lower amount of genetic diversity (along with other populations of the North West; see Salas *et al.* 1998) than other populations of the Iberian peninsula. Probably, the posterior arrival of the Romans did not significantly alter the genetic landscape of the Iberian peninsula; beyond the initial military conquest, it was probably more an acculturation phenomenon than a migration movement. Later arrivals of Germanic peoples into the Iberian peninsula, such as the Visigoths (5th century AD) or the Vandals (6th century AD) were probably too small to significantly alter the existing Roman human substratum. Nonetheless, the accumulative impact of these migrations could explain the increased genetic diversity found in present-day Iberian populations.

A putative ancestral relationship between Iberians and Etruscans, as representatives of the general pre-Roman European substratum, is also not supported by our results. The Iberians are not especially close to present populations from Central Italy; moreover, few haplotypes seem to be shared between the ancient Iberians and the contemporary Etruscan sequences, retrieved by Vernesi *et al.* (2004). Most probably, haplotypes labelled as 6AM (16,126 substitution, either *MseI* 14,766+ or *MseI* 14,766-) correspond to the 16,126 haplotype from the H and J haplogroups, respectively, found in the Iberians. Apart from the CRS other haplotypes are similar, but not exactly the same, as those found in the Etruscans.

Contrary to the continuity we observed in the Iberians, there is limited genealogical continuity between the Etruscans and their modern counterparts from Tuscany and Central Italy; this points to a high rate of extinction of mitochondrial haplotypes, at least in the Etruscans. Vernesi *et al.* (2004) suggested that many lineages described in the Etruscans have been lost by drift

or historical processes. Most Etruscan tombs discovered belonged to the social elite, and thus, the fact that few Etruscan sequences have an exact match in the modern database may reflect social factors related to the Roman assimilation of the Etruscan society.

By contrast, almost all the Iberian sequences have been found in modern European populations, so in the Iberian peninsula the process seems to have been the contrary: an enrichment of genetic diversity, but from a source having a similar haplogroup composition (albeit more diverse) to that of Iberians. This could be related to the fact that the skeletal remains studied did not belong to a social elite, because the Iberian people were routinely cremated. In fact our sample, although probably biased, might fairly well represent the genetic substratum of the normal Iberian population. It is likely that the newly arrived Romans mainly become admixed with the indigenous people, after routing the Iberian military elite (Plácido, 1998). Moreover, we hypothesize that the limited genetic diversity of the ancient Iberians could be due to the existence of small populations and endogamic processes related to the tribe-structured Iberian society (Ruiz & Molinos, 1998). However, we cannot discard the possibility that restricted geographical sampling is biasing our results.

Along with the Etruscans, this is the only study where a distinctive pre-Roman population has been genetically studied, following the strictest authentication criteria for ancient DNA. Our data suggest that long-term genetic continuity has existed in the Iberian peninsula since at least the 6th century BC to present times. The documented, posterior arrivals of groups from Europe and North Africa did not significantly alter the pre-Roman genetic background; however, they probably increased the relatively low genetic diversity of the Iberian groups.

Acknowledgments

We are grateful to Aurora Martín and Enriqueta Pons (Museu d'Arqueologia de Catalunya, Girona) for granting us access to the Iberian samples. Roger Anglada (UPF), Oscar Fornas (UPF), Mónica Vallès (UPF) and Nuria Naveran (Universidad Santiago de Compostela) provided technical support in the quantitation analysis. This research was supported by the Dirección General de Investigación, Ministerio de Ciencia y Tecnología of Spain (grants BOS2001-0794 and

BMC2001-0772), by the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (grant 2001SGR00285), by the I.E.C. (Institut d'Estudis Catalans) and a fellowship to M.L.S. (AP2002-1065).

References

- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J. -M., Roostalu, U., Loogväli, E. -L., Kivisild, T., Bandelt, H. -J., Richards, M., Villems, R., Santachiara-Benerecetti, A. S., Semino, O. & Torroni, A. (2004) The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. *Am. J. Hum. Genet.* **75**, 910–918.
- Adroher, A. M., Pons, E., Ruiz de Arbullo, J. (1993) El yacimiento de Mas Castellar de Pontós y el comercio del cereal ibérico en la zona de Emporion y Rhode (ss. IV-II aC). *Archivo Español de Arqueología* **66**, 31–70.
- Agustí, B. (1999) Valoració del material dentari en poblats ibèrics empordanesos. In : *Els productes alimentaris d'origen vegetal a l'Edat del Ferro de l'Europa Occidental: de la producció al consum*. Museu d'Arqueologia de Catalunya (eds. Buxó, R. & Pons, E.) Girona. pp. 403–408.
- Alonso, A., Martín, P., Albarrán, C., Garcia, P., Garcia, O., Fernandez 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. *For. Sci. Int.* **139**, 141–149.
- Anderson, S., Bankier, A. T., Barrel, B. G., De Bruijn, M.H.L., 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.
- Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753.
- Bertranpetit, J. & Cavalli-Sforza, L. L. (1991) A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann. Hum. Genet.* **55**, 51–67.
- Bertranpetit, J., Sala, J., Calafell, F., Underhill, P. A., Moral, P. & Comas, D. (1995) Human mitochondrial DNA variation and the origin of Basques. *Ann. Hum. Genet.* **59**, 63–81.
- Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A. & Bertranpetit, J. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* **68**, 1019–1029.
- 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.
- Cooper, A.R. & Poinar, H. (2000) Ancient DNA: do it right or not at all. *Science* **289**, 1139.
- Corte-Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Isaad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B. (1996) Genetic diversity in the Iberian Peninsula determined for mitochondrial sequence analysis. *Ann. Hum. Genet.* **60**, 331–350.
- Crespillo, M., Luque, J. A., Paredes, M., Fernández, R., Ramírez, E. & Valverde, J. L. (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. *Int. J. Legal Med.* **114**, 130–132.
- Francalacci, P., Bertranpetit, J., Calafell, F. & Underhill, P. A. (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am. J. Phys. Anthropol.* **100**, 443–460.
- Gilbert, M.T.P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N. & Cooper, A. (2003) Distribution patterns of Postmortem damage in human mitochondrial DNA. *Am. J. Hum. Genet.* **72**, 32–47.
- Handt, O., Krings, M., Ward, R. H. & Pääbo, S. (1996) The retrieval of ancient human DNA sequences. *Am. J. Hum. Genet.* **59**, 368–376.
- Harrison, R. J. (1988) *Spain at the dawn of history*. Thames and Hudson. London.
- Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A. & Pääbo, S. (2001) DNA sequences from multiple amplifications reveal artefacts induced by cytosine deamination in ancient DNA. *Nucl. Acid Res.* **29** (23), 4793–4799.
- Hoz, J. de (1995) El poblamiento antiguo de los Pirineos desde el punto de vista lingüístico” In: *Muntanyes i població. El passat dels Pirineus des d’una perspectiva multidisciplinària* (eds. Bertranpetit, J. & Vives, E.) Centre de Trobada de les Cultures Pirinenques, Andorra la Vella, pp: 271–299.
- 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.
- López-Soto, M. & Sanz, P. (2000) Mitochondrial DNA polymorphisms in individuals living in Andalucía (south of Spain) and Extremadura (western Spain). *Cuadernos de Medicina Forense*, 17–24.
- Martín, M. A. (1985) *Ullastret. Poblal ibèric*. Diputació de Girona i Departament de Cultura de la Generalitat de Catalunya.
- Pereira, L., Prata, M. J. & Amorim, A. (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann. Hum. Genet.* **64**, 491–506.
- Plácido, D. (1998) Las sociedades mediterráneas y el imperio romano: diversidad e integración de los sistemas económicos. In: *Transiciones en la antigüedad y feudalismo* (Ed. J. Trías) Madrid, FIM, pp. 9–23.
- Poinar, H. N., Hoss, M., Bada, J. L. & Pääbo, S. (1996) Amino acid racemization and the preservation of ancient DNA. *Science* **272**, 864–866.
- Poinar, H. N. (2004) Criteria for authenticity in ancient DNA work. In: *Human Evolutionary Genetics* (eds. M. A. Jobling, M. E. Hurles and C. Tyler-Smith) Taylor and Francis Group, New York. Pp: 115–116.
- 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.
- Rando, J. C., Pinto, F., Gonzalez, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H. J. (1998) Mitochondrial DNA análisis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann. Hum. Genet.* **62**, 531–550.
- Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, 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., Golge, 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-Zaheri, N., Santachiara-Benerecetti, S., Scozari, 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.
- Ruiz, A. & Molinos, M. (1998) *The Archaeology of the Iberians*. Cambridge University Press, Cambridge.
- Salas, A., Comas, D., Lareu, M. V., Bertranpetit, J. & Carracedo, A. (1998) MtDNA analysis of the Galician population: a genetic edge of European variation. *Eur. J. Hum. Genet.* **6**, 365–375.
- Schneider, S., Roessli, D. & Excoffier, L. (2000) Arlequin ver. 2.000: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- 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

- Stone, A. C. & Stoneking, M. (1998) MtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am. J. Hum. Genet.* **62**, 1153–1170.
- Tagliabracci, A., Turchi, C., Buscemi, L. & Sassaroli, C. (2001) Polymorphism of the mitochondrial DNA control region in Italians. *Int. J. Legal Med.* **14**, 224–228.
- Torroni, A., Bandelt, H. J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonne-Tamir, B. & Scozzari, R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* **62** (5), 1137–1152.
- Vernesi, C., Di Benedetto, G., Caramelli, D., Secchieri, E., Katti, E., Malaspina, P., Novelletto, A., Terribile Wiel Marin, A. & Barbujani, G. (2001) Genetic characterization of the body attributed to the evangelist Luke. *Proc. Natl. Acad. Sci. USA* **98**, 13460–13463.
- Vernesi, C., Caramelli, D., Dupanloup, I., Bertorelle, G., Lari, M., Cappellini, E., Moggi-Cecchi, J., Chiarelli, B., Castri, L., Casoli, A., Mallegni, F., Lalueza-Fox, C. & Barbujani, G. (2004) The Etruscans: a population genetic study. *Am. J. Hum. Genet.* **74** (4), 694–704.
- Wakeley, J. (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J. Mol. Evol.* **37** (6), 613–623.

Supplementary Material

Table 1: Sequences, primers and clones used to reconstruct the HVR1 sequences from 17 Iberians (S: sample number as in Tables 3 and 4). Dots indicate identity to the reference sequence, substitutions are displayed. I: independent replication; UNG: samples treated with UNG enzyme; E: second extraction from the same individual. Different amplifications of the same fragment display the primers at the extremes of the sequences; clones generated from a particular PCR product are listed below the corresponding direct sequence.

Received: 24 August 2004

Accepted: 9 December 2004