

Joining the Pillars of Hercules: mtDNA Sequences Show Multidirectional Gene Flow in the Western Mediterranean

S. Plaza¹, F. Calafell¹, A. Helal², N. Bouzerna³, G. Lefranc⁴, J. Bertranpetit¹ and D. Comas¹

¹Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona

²Laboratoire d'Immunogénétique, Faculté de Pharmacie, Monastir, Tunisia

³Laboratoire de Biochimie et Microbiologie Appliquée, Unité d'Immunobiologie, Université d'Annaba, Annaba, Algérie

⁴Institut de Génétique Humaine, UPR CNRS 1142, et Université Montpellier II, 34095 Montpellier Cedex 5, France

Summary

Phylogenetic analysis of mitochondrial DNA (mtDNA) performed in Western Mediterranean populations has shown that both shores share a common set of mtDNA haplogroups already found in Europe and the Middle East. Principal co-ordinates of genetic distances and principal components analyses based on the haplotype frequencies show that the main genetic difference is attributed to the higher frequency of sub-Saharan L haplogroups in NW Africa, showing some gene flow across the Sahara desert, with a major impact in the southern populations of NW Africa. The AMOVA demonstrates that SW European populations are highly homogeneous whereas NW African populations display a more heterogeneous genetic pattern, due to an east-west differentiation as a result of gene flow coming from the East. Despite the shared haplogroups found in both areas, the European V and the NW African U6 haplogroups reveal the traces of the Mediterranean Sea permeability to female migrations, and allowed for determination and quantification of the genetic contribution of both shores to the genetic landscape of the geographic area.

Comparison of mtDNA data with autosomal markers and Y-chromosome lineages, analysed in the same populations, shows a congruent pattern, although female-mediated gene flow seems to have been more intense than male-mediated gene flow.

Introduction

The western Mediterranean populations have experienced a long, intricated history that, too often, has been considered separately for the African and European shores, or from an exclusively European perspective. Both the African and the European shores have acted as termini of population expansions. The independent and parallel colonisation from the East of both areas by anatomically modern humans in Palae-

olithic times, and the expansion of farming during the Neolithic, have modelled the genetic landscape of both areas. Moreover other demographic events, such as the expansion of the Arabisation along the Maghrib, have also come from the East arriving in NW Africa.

Genetic diversity studies have provided a major insight into human evolution on a global scale, but they have also been useful in regional studies. Population processes such as expansions, migrations, dispersals and admixtures leave a footprint in the genetic composition of the groups that allow us to trace back population history. Several genetic markers have been analysed in the westernmost part of the Mediterranean in order to extricate such processes. The compilation of classical genetic markers (Bosch *et al.* 1997; Simoni *et al.* 1999)

*Correspondence: David Comas, Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Spain. Tel: +34 93 542 28 02; Fax: +34 93 542 28 44. E-mail: david.comas@cexs.upf.es

has shown a clear genetic differentiation between the northern and southern coasts, attributed to independent parallel expansions along the two shores followed by little gene flow across the Mediterranean. Nevertheless, there is some contradictory data, based on HLA polymorphisms, on the degree of genetic relationship between both coasts in West Mediterranean populations (Arnaiz-Villena *et al.* 1995; Comas *et al.* 1998). Analyses of autosomal STRs (Bosch *et al.* 2000) and Alu insertion polymorphisms (Comas *et al.* 2000) confirmed the genetic difference between both groups of populations, also detecting some Sub-Saharan genetic flow into NW African populations. The high-resolution analysis of Y-chromosome biallelic and STR markers (Bosch *et al.* 2001) has revealed clear genetic differentiation due to a major independent Upper Palaeolithic contribution in both areas, followed by gene flow from the Near East during the Neolithic, and small bidirectional gene flow across the Mediterranean. Several mitochondrial DNA (mtDNA) analyses have focused in the structure of Iberian populations (Bertranpetit *et al.* 1995; C  rte-Real *et al.* 1996; Salas *et al.* 1998; Pereira *et al.* 2000), of NW African populations (Rando *et al.* 1998; Brakez *et al.* 2001), and their relation to the Canary Islands (Pinto *et al.* 1996). Nevertheless, no analysis has jointly considered the population relationships of Western Mediterranean populations using mtDNA sequences.

The analysis of mitochondrial DNA diversity has been one of the most successful tools applied to unravel regional population histories. Two different approaches have been followed in order to perform mtDNA analyses: the sequencing of the hypervariable segments of the non-coding part of the molecule, the control region, and the study of the coding region through high-resolution RFLPs. The joint analysis of both kinds of markers (control region sequences and RFLPs in the coding region) has proven to be a powerful tool in studying human diversity (Torrioni *et al.* 1996), and has led to the construction of robust phylogenies of mtDNA sequences (Macaulay *et al.* 1999), which allow one to elucidate human demographic scenarios.

In the present study, we have analysed the hypervariable segment I (HVSI) of the control region in several Western Mediterranean populations, and have added the information yielded by three SNPs in the mtDNA-

coding region in order to ascribe the mtDNA variation to specific branches of the gene genealogy. This analysis allows us to describe the genetic landscape of the geographic region, compare it to that obtained with other genomic regions (particularly those with a clear phylogeography, such as the Y-chromosome), and interpret it in terms of external gene flow and of exchanges between the northern and southern shores of the Mediterranean.

Material and Methods

Population Samples

A data set comprising sequences for the first hypervariable segment (HVSI) of the mtDNA control region (positions 16024 to 16383, according to the Cambridge Reference Sequence; CRS, Anderson *et al.* 1981; Andrews *et al.* 1999) in populations of the Western Mediterranean (defined as the Iberian and Italian Peninsulas and intervening islands, southern France, and NW Africa from Tunisia to Mauritania), was collected from the literature and from our own analyses. Sequences for a total of 1,719 individuals were collected (see populations, sample sizes and references in Table 1 and Figure 1); of those, we sequenced a total of 267 unrelated individuals: 172 from North-West Africa and 95 from the Iberian Peninsula. The NW African samples included 56 Saharawi, 18 Moroccans Arabs, 4 Berbers from North-Central Morocco, 47 Algerians and 47 Tunisians. The Iberian Peninsula samples comprised 49 Andalusians and 46 Catalans. Populations analysed were chosen in order to generate a complete picture of the region. These sequences are available at <http://www.upf.es/cexs/recerca/bioevo/index.htm>

MtDNA Amplification and Sequencing

Total DNA was extracted from fresh blood using standard phenol-chloroform methods after digestion with proteinase K. HVSI was amplified with primers and methods as described elsewhere (Mateu *et al.* 1997). The amplified product was purified with the Gene Clean kit (BIO 101) and sequencing reactions were performed using the Big Dye Terminator (version 3.0) Cycle Sequencing Kit, with AmpliTaq[®] DNA Polymerase (Applied Biosystems). The sequencing products were

Table 1 Diversity parameters for the HVRI in several populations

Populations	n	k	S	Sequence diversity	Mean pairwise differences	Nucleotide diversity	References
Northwest Africa	429						
Algerians	47	27	51	0.957±0.043	5.72±4.28	0.0158	1
Mauritanians	30	23	31	0.975±0.025	6.09±3.91	0.0169	2
Moroccan Arabs	50	44	68	0.993±0.007	7.04±2.96	0.0195	1,2
Moroccan Berbers	64	42	51	0.968±0.032	4.52±5.48	0.0125	1,2
Mozabites	85	29	35	0.942±0.058	4.73±5.27	0.0131	3
Saharawi	56	41	46	0.978±0.022	5.45±4.55	0.0151	1
Southern Berbers	50	34	38	0.941±0.059	4.60±5.40	0.0128	4
Tunisians	47	42	61	0.989±0.010	6.15±3.85	0.0171	1
Iberian Peninsula	887						
Andalusians	158	106	82	0.965±0.035	4.26±5.74	0.0118	1,5,6,7,8
Basques	173	71	64	0.942±0.058	3.03±6.97	0.0084	5,8,9,10
Catalans	78	48	45	0.938±0.062	3.66±6.34	0.0102	1,5,6,8
Central Spain	50	38	49	0.953±0.047	4.59±5.41	0.0128	6,8
Galicians	103	62	61	0.939±0.061	3.31±6.69	0.0092	6,8,11
Valencians	30	24	37	0.970±0.030	4.25±5.75	0.0118	8
Portuguese	54	38	40	0.934±0.066	3.60±6.40	0.0100	5
Nportuguese	100	67	71	0.953±0.047	4.78±5.22	0.0133	12
Cportuguese	82	62	66	0.977±0.023	4.87±4.13	0.0135	12
Sportuguese	59	41	55	0.943±0.057	4.54±5.46	0.0126	12
Italy	411						
Central Italy	83	63	61	0.974±0.012	4.78±5.41	0.0133	13
Sardinians	73	50	57	0.955±0.045	4.24±5.76	0.0118	14
Sicilians	169	97	92	0.936±0.064	4.03±5.97	0.0112	15,16
Southern Italy	37	31	47	0.969±0.031	4.86±5.14	0.0135	16
Tuscans	49	40	55	0.969±0.031	5.03±4.97	0.0140	17

n: number of individuals; k: number of different sequences; S: number of variable positions. NPortuguese: Northern Portuguese; CPortuguese: Central Portuguese; SPortuguese: Southern Portuguese. References: 1 Present study; 2 Rando *et al.* 1998; 3 Macaulay *et al.* 1999; 4 Brakez *et al.* 2000; 5 Côrte-Real *et al.* 1996; 6 Crespillo *et al.* 2000; 7 López-Soto *et al.* 2000; 8 A. Alonso (personal communication); 9 Bertranpetit *et al.* 1995; 10 Richards *et al.* 2000; 11 Salas *et al.* 1998; 12 Pereira *et al.* 2000; 13 Tagliabracci *et al.* 2001; 14 Di Rienzo & Wilson, 1991; O. Rickards (personal communication); 15 Cali *et al.* 2001; 16 O. Rickards *et al.* 2000 and personal communication; 17 Francalacci *et al.* 1996.

run in an ABI PRISM 3100 sequencer (Applied Biosystems).

Three positions in the mtDNA coding region (10400, 12308 and 12705, according to Anderson *et al.* 1981) were also determined in some of the sequenced individuals by using the SNaPshotTM ddNTP Primer Extension Kit (Applied Biosystems), as described elsewhere (Comas *et al.* in preparation), which implements a single-base primer extension protocol that uses labelled ddNTPs to interrogate SNPs.

Phylogenetic Analysis

Sequence alignment was performed using the ESEE program (Cabot, 1988). Each control region sequence

was assigned to a given haplogroup by comparison with the data sets where mtDNA had been typed for both RFLPs and HVSI sequences (Torroni *et al.* 1996; Watson *et al.* 1997; Rando *et al.* 1998; Macaulay *et al.* 1999) and the data sets of Richards *et al.* (2000), and their classification scheme was used with a single modification: a few sequences bearing a transition at position 16126 and not carrying transitions at 16069, 16294, 16296, or 16362, were classified as J/T, a denomination that should not necessarily imply that they are in a group ancestral to haplogroups J and T. When the information given by the nucleotide substitutions of the HVSI in those individuals sequenced in the present study was insufficient to assign a sequence to a given haplogroup, three positions outside the control region (10400, 12308

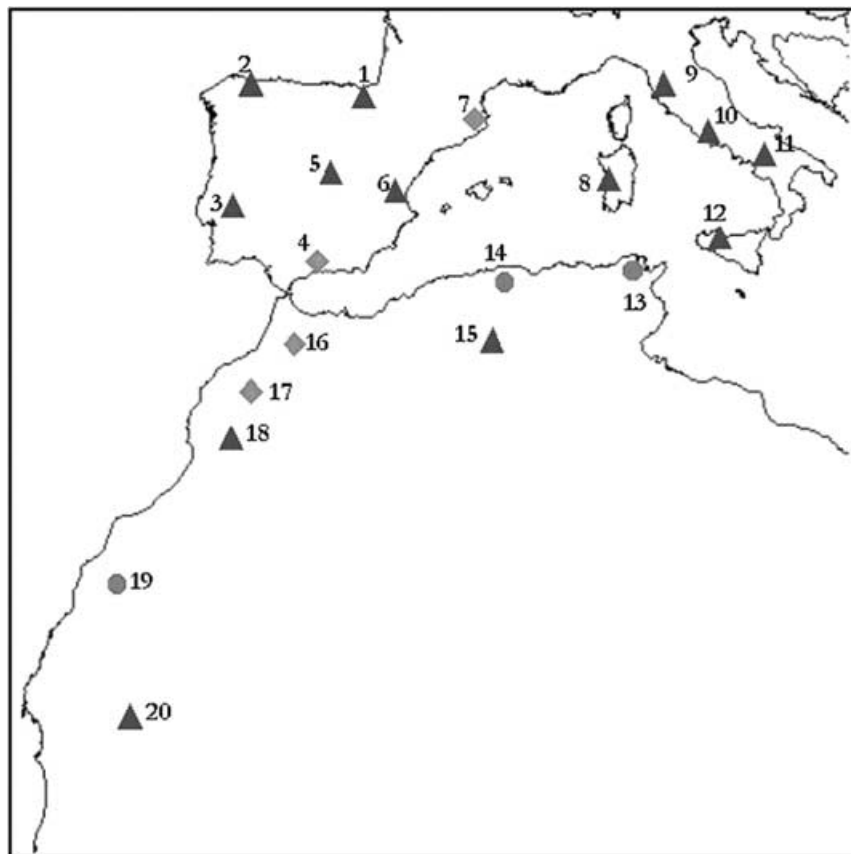


Figure 1 The Western Mediterranean. Geographical location of the samples analysed: 1: Galicians; 2: Basques; 3: Portuguese (include mixed Portuguese, Northern, Central and Southern Portuguese); 4: Andalusians; 5: Central Spain; 6: Valencians; 7: Catalans; 8: Sardinians; 9: Tuscans; 10: Central Italy; 11: Southern Italy; 12: Sicilians; 13: Tunisians; 14: Algerians; 15: Mozabites; 16: Moroccan Berbers; 17: Moroccan Arabs; 18: Southern Berbers; 19: Saharawi; 20: Mauritians. Dots represent samples sequenced in the present work, squares represent samples pooled from the literature as well as new sequences included, and triangles represent samples taken from the literature.

and 12705) were determined. These positions allowed us to assign the control region sequences to three different major haplogroups: 10400T defines the major M haplogroup, 12308G defines the major U haplogroup (including the K haplogroup), and 12705C defines the major R haplogroup, which includes a large set of haplogroups (H, V, J, T, U, B and F). Nevertheless, 2% of all sequences remained ambiguous or could not be typed for these three positions as they were taken from the literature, and they were classified as “other”.

The networks relating HVSI sequences within some of the haplogroups described were constructed by using a reduced-median algorithm (Bandelt *et al.* 1995) as implemented in the Network 3.0 program. The dating method employed (Morral *et al.* 1994; Saillard *et al.*

2000) is based on the average number of mutations accumulated from an ancestral sequence as a linear function of time and mutation rate. This method was also performed with the Network 3.0 program.

Population Analysis

Population internal genetic diversity parameters (nucleotide diversity, sequence diversity and mean pairwise differences) were computed with the Arlequin 2000 program (Schneider *et al.* 1996).

Population genetic structure was tested through analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992), using the Arlequin 2000. Genetic distances between populations using the first mtDNA hypervariable region were calculated by intermatch-mismatch

pairwise differences according to the equation $D = d_{ij} - (d_{ii} + d_{jj})/2$ (Nei 1987), where d_{ij} is the mean pairwise differences between populations i and j , and d_{ii} and d_{jj} are the mean pairwise differences within populations i and j respectively. The distance standard errors were computed by resampling nucleotide positions with 1,000 bootstrap iterations (Efron, 1982). A principal co-ordinate plot (Gower, 1966) was also obtained from the distance matrix. Principal component analysis was performed from haplotype frequencies using the SPSS package.

Results

Phylogeographic Structure

Haplogroup frequencies estimated as described above are listed in Table 2. The phylogeographic structure of mtDNA in the Western Mediterranean can be summarised as five sets of haplogroups: 1) sub-Saharan haplogroup L (including L1, L2, L3); 2) haplogroups J, T, J/T; 3) haplogroups H, V, HV; 4) haplogroup U (including K); and 5) haplogroups W, I, X, and M.

L haplogroups are relatively infrequent in Italians (with a maximum of 8.1% in South Italians) and Iberians (with a maximum of 6.1% in Central Portuguese). On the contrary, L haplogroups are distributed in all North African populations at high frequencies (from 26% in South Berbers to 43.5% in Mauritians) with the exception of Mozabites (12.9%) and Moroccan Berbers (3.2%). In fact, the frequency of the L haplogroups in Moroccan Berbers is similar to that found in Iberians and Italians. The frequency of the L haplogroups might represent the sub-Saharan genetic flow into the populations analysed, which has shown to be substantial in NW Africa but very limited in European populations.

In the populations analysed, haplogroups J and T present their highest frequencies in the Italian samples, with values over 15%. Iberians showed a heterogeneous frequency distribution with values that range from 6.6% in Valencians to 18.7% in Southern Portuguese. NW Africans have similar J and T frequencies to Europeans, although it is worth noting that Saharawi and Mauritians, the southern NW African samples, differ from the other populations in that haplogroups J and T are almost absent.

Haplogroup U is found in all samples analysed at considerable frequencies. The most relevant aspect within this group of sequences is the presence of haplogroup U6, to which a North African origin has been attributed (Rando *et al.* 1998). Haplogroup U6 is largely distributed among Mozabites (28.2%) and Mauritians (20%). In other NW Africans, the frequency of U6 ranges from 4.2% in Tunisians to 8% in Moroccan Arabs, with the remarkable case of Algerians where haplogroup U6 is absent. In Italians, haplogroup U6 is practically absent, with only one sequence found among Sicilians. In the Iberian Peninsula U6 distribution is sparse. It is present in the south-western part of the Peninsula at low frequencies (<7%), and is absent in Basques, Catalans, Valencians, Central Portuguese, and Southern Portuguese. Few U6 sequences are found in other populations from different geographical regions: Sub-Saharan and NE Africa, the Middle East and the Canary Islands (with a frequency of 14%; Rando *et al.* 1998, 1999). The structure of the variation of U6 sequences is shown in Figure 2, from which the age of U6 can be estimated at $47,000 \pm 18,000$ years, similar to that first estimated by Rando *et al.* (1998). The network shows a clear structure in subhaplogroups within U6: U6a (characterised by 16278T; Rando *et al.* 1999), U6a1 (characterised by 16278T and 16189C; Richards *et al.* 2000), and U6b (characterised by 16311C; Rando *et al.* 1999). The present Iberian and NW African sequences are found within haplogroups U6a and U6a1, but haplogroup U6b contains no NW African sequences and is mainly composed of Canarian and Iberian sequences. Moreover, there is a group of sequences within U6b characterised by 16163T, which we named U6b1 (although unnamed, this was already discussed by Rando *et al.* 1999), which presents basically Canarian sequences. The age of this group of sequences (Canarians plus two Iberian sequences) is around $9,400 \pm 5,500$ years. The presence of U6a and U6a1 haplogroups in the Iberian Peninsula could be attributed to gene flow from NW Africa, and the most plausible origin for U6b1 lineages in Iberia is recent gene flow from the Canary Islands after the contact between Europeans and the Canary aborigines in the fifteenth century.

H and V represent the major group of sequences in Iberia and Italy. H is by far the most frequent haplogroup in western European populations, as it is in all of Europe

Table 2 Haplogroup frequencies (%) in Western Mediterranean Populations

Pop (n)	Alg (47)	Mau (30)	MA (50)	MB (64)	MB (64)	Moz (85)	Moz (85)	Sah (56)	SBer (50)	Tun (47)	NWA (429)	And (158)	Bas (173)	Cat (78)	CS (50)	Gal (103)	Val (30)	Port (54)	NPo (100)	CPo (82)	SPo (59)	IBE (887)	CIt (83)	Sard (73)	Sic (169)	SIt (37)	Tus (49)	ITA (411)
L1	6.4	23.4	10.0	1.6	-	3.6	6.0	-	6.4	-	-	-	-	-	2.0	-	-	-	1.0	1.3	1.7	0.6	-	1.4	-	2.7	-	0.8
L2	6.4	13.4	6.0	-	5.9	7.1	10.0	12.8	7.7	0.6	-	0.6	-	-	2.0	1.0	3.4	1.8	3.0	2.4	1.7	1.6	-	1.4	0.6	-	-	0.4
L3	14.9	6.7	16.0	1.6	7.0	23.3	10.0	14.9	11.8	1.3	-	1.3	-	1.3	-	1.0	-	1.8	1.0	2.4	1.7	1.0	1.2	-	-	5.4	2.0	1.7
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.2	-	-	-	-	0.2
M1	12.8	-	2.0	-	4.7	-	-	-	4.2	3.0	1.9	-	-	-	-	-	-	-	-	1.3	-	0.3	-	1.4	1.8	-	-	0.6
M5	-	-	-	-	-	-	-	-	-	-	1.9	-	1.9	-	2.0	-	-	-	-	-	-	0.4	-	-	-	-	-	-
N	-	-	-	-	-	-	-	-	-	0.6	-	0.6	-	1.3	-	-	6.6	-	-	1.3	-	1.0	-	1.4	1.2	2.7	-	1.1
I	-	-	-	-	-	-	-	-	2.1	0.3	1.9	-	-	-	-	-	-	-	1.0	-	1.7	0.5	-	1.2	-	4.1	1.1	
W	-	-	-	-	-	-	-	-	-	1.3	-	1.3	-	5.1	2.0	1.9	-	-	2.0	1.3	-	1.4	1.2	1.4	1.8	5.4	2.0	2.4
X	2.1	-	4.0	-	-	-	-	-	2.1	1.0	3.2	1.7	2.6	2.0	1.0	-	-	-	-	3.6	1.7	1.6	3.6	1.4	2.9	5.4	6.1	3.9
R1	-	-	-	-	-	-	-	-	-	-	1.7	-	1.7	-	-	-	-	-	-	-	-	0.2	-	-	-	2.7	-	0.5
J/T	-	-	-	1.6	-	-	-	-	-	0.2	0.6	-	0.6	-	-	-	-	-	-	-	-	0.1	-	-	1.8	2.7	-	0.9
T	4.2	-	4.0	15.6	4.7	1.8	4.0	6.4	5.1	4.4	5.2	7.7	10.0	2.9	-	-	-	11.1	11.0	11.0	10.2	7.3	15.7	12.3	8.3	13.5	10.2	12.0
J	12.8	3.3	4.0	9.4	3.5	-	10.0	4.2	5.9	7.0	4.0	2.6	8.0	9.7	6.6	5.6	6.0	6.1	8.5	6.1	8.5	6.5	6.0	5.5	5.3	2.7	14.3	6.8
U*	2.1	3.3	12.0	6.2	12.9	8.9	8.0	6.4	7.5	10.1	13.3	9.0	20.0	12.6	16.6	9.3	16.0	14.6	11.9	13.1	9.6	12.3	4.7	5.4	12.2	8.8	-	-
U6	-	20.0	8.0	7.8	28.2	5.4	6.0	4.2	9.9	1.9	-	2.0	1.9	-	2.0	1.9	-	5.6	7.0	-	-	1.8	-	-	0.6	-	-	0.1
K	4.2	6.6	4.0	7.8	-	7.1	2.0	6.4	4.8	6.3	5.2	6.4	2.0	3.9	10.0	7.4	3.0	7.3	6.8	5.8	7.2	5.8	7.2	5.5	2.9	2.7	8.2	5.3
HV	-	-	-	-	-	-	7.1	2.0	6.4	1.9	2.5	-	1.3	2.0	1.0	3.4	3.7	-	-	1.4	1.2	2.7	-	1.2	2.7	-	2.0	1.2
H	34.0	20.0	26.0	42.2	24.7	17.9	32.0	23.4	27.5	46.2	57.8	56.4	46.0	59.2	53.3	48.1	41.0	37.8	44.1	49.2	47.0	50.7	50.3	45.9	38.8	46.5	-	-
V	-	3.3	4.0	6.2	8.2	17.9	10.0	-	6.2	5.7	10.4	5.1	-	2.9	-	3.7	8.0	7.3	6.8	5.0	4.8	2.7	5.9	2.7	5.9	2.7	-	3.2
Other	-	-	-	-	-	-	-	6.4	0.8	2.5	0.6	1.3	-	1.0	-	2.4	3.4	1.3	2.4	3.4	1.3	1.2	-	10.6	-	-	-	2.9

Alg: Algerians; Mau: Mauritians; MA: Moroccan Arabs; MB: Moroccan Berbers; Moz: Mozabites; Sah: Saharawis; SBer: South Berbers; Tun: Tunisians; And: Andalusians; Bas: Basques; Cat: Catalans; CS: Central Spain; Gal: Galicians; Val: Valencians; Port: Portuguese; NPo: North Portuguese; CPo: Central Portuguese; SPo: South Portuguese; CIt: Central Italy; Sard: Sardinians; Sic: Sicilians; SIt: South Italians; Tus: Tuscans. NWA: unweighted average frequencies in NW Africans; IBE: unweighted average frequencies in Iberians; ITA: unweighted average frequencies in Italians. (*): Excluding U6. N includes sequences carrying the HVRI substitutions diagnostic of either N1a or N1b.

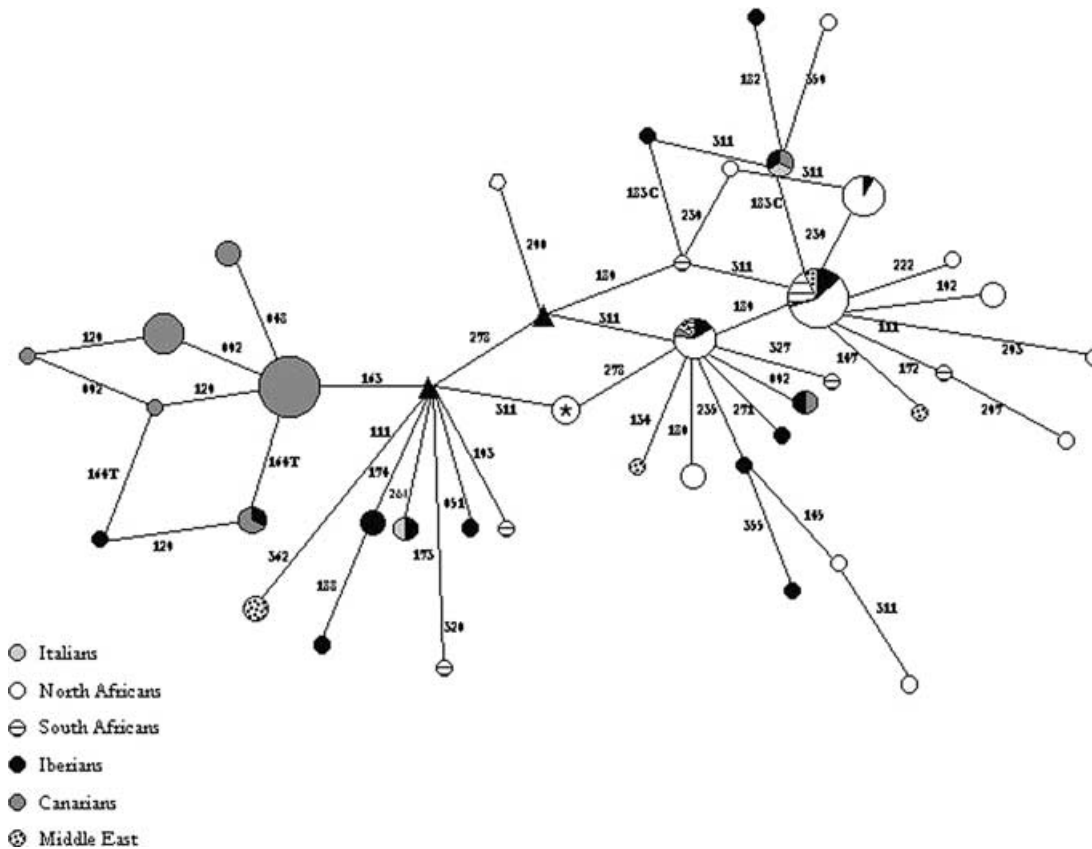


Figure 2 Phylogenetic network of haplogroup U6 HVRI sequences (present data and data from Richards *et al.* 2000). The size of the circles is proportional to the number of sequences. The node marked with an asterisk indicates the ancestral sequence (16172C, 16219G). Numbers along links refer to nucleotide positions in HVRI minus 16000; suffixes indicate a transversion. Subhaplogroups U6a and U6a1, to the right in the graph, are defined by positions 16278T and 16278T, 16189C respectively. Subhaplogroups U6b and U6b1, to the left in the graph, are defined by positions 16311C and 16163T, 16311C respectively.

(Simoni *et al.* 2000a; Richards *et al.* 2000). It has been suggested that haplogroup V originated and expanded from NE Iberia (Torroni *et al.* 1998; Torroni *et al.* 2001). In the European samples analysed, its frequency (which includes pre-V and V proper as defined by Torroni *et al.* 2001) ranges from 2.7% in Sardinia and Southern Italy to 10.4% in Basques, and is absent in Central Spaniards, Valencians, and Tuscans. Except in Algerians and Tunisians, haplogroup V has been found in all the samples analysed, with high frequencies among the Saharawi (17.9%) and Southern Berbers (10%). In order to elucidate the phylogenetic relationships between sequences, a network of V sequences was constructed (Figure 3). The network displayed a clear star-like pattern with all V sequences found in NW Africa close to the V sequence root type or with one or two added

substitutions, whereas Italian and Iberian V sequences show a wider distribution of substitutions. Out of the five different V haplotypes found in NW Africa, three were those that are most frequent in Europe, while only two were specific to NW Africa. A time depth for the haplogroup V of $13,700 \pm 3,000$ years was estimated when all sequences were included, similar to previous estimates (Torroni *et al.* 2001).

The last section of the mtDNA phylogeny considered includes the Eurasian haplogroups W, I, X, and haplogroup M. Haplogroups W, I, and X are basically found in continental Italy, and some traces are found in Iberians, Algerians, Tunisians and Moroccan Arabs. The M sequences found in the analysed populations can be sorted into two different phylogenetic groups: haplogroups M1 and M5. It has been suggested that

Table 3 Analyses of Molecular Variance (AMOVA) in West Mediterranean populations

Groups	Among groups		Among populations within groups		Within populations	
	with L	without L	with L	without L	with L	without L
All populations			2.56**	2.00**	97.44**	98.00**
NW Africans			3.53**	4.50**	96.47**	95.50**
SW Europeans			0.62**	0.62**	99.38**	99.38**
NW Africa vs SW Europe	2.62**	1.41**	1.48**	1.49**	95.90**	97.10**
Iberian Peninsula vs Italy	0.06 ns	0.11 ns	0.59**	0.57**	99.34**	99.32**
Eastern vs Western NW Africa ^a	0.94 ns	1.96*	1.29**	0.69 ns	97.76**	97.35**

** ($p < 0.01$); * ($p < 0.05$); ns : non-significant All the analyses were performed taking into account lineages belonging to L haplogroups (with L) and ignoring L lineages (without L).

^aTwo groups: Algerians and Tunisians versus the rest of NW African populations.

L lineages were removed. Considering separately the southern and northern populations, NW Africans are more heterogeneous: Φ_{ST} among NW Africans is 3.5% (4.5% without L sequences), as compared to 0.6% among SW Europeans.

When we grouped the samples according to their geographical area (SW Europeans versus NW Africans), 1.5% of the genetic variance was due to differences between samples of the same geographical area, and 2.6% was attributable to differences between geographical areas. The variance attributable to differences among geographical groups decreased to 1.4% when the L sequences were removed, whereas the variance attributable to differences within groups did not vary, showing that Sub-Saharan gene flow into NW Africa has in part been responsible for the differences between the two groups.

In order to establish a valid comparison between nuclear DNA, Y chromosomal, and mtDNA, we performed an AMOVA with Alu polymorphisms (Comas *et al.* 2000), the Y-chromosome lineages (Bosch *et al.* 2001) and mtDNA data among the same populations from NW Africa and the Iberia Peninsula as described in Bosch *et al.* (2001). We found that the proportion of the genetic variance that can be accounted for between the NW African and Iberian populations for mtDNA is 0.86% ($p = 0.053$), 1.89% ($p = 0.028$) for Alu insertion polymorphisms, and 35.2% ($p = 0.024$) for the Y chromosome. It is not surprising to find that the results show clear differences between male and female lineages due to the already described sexual differential migration

patterns for worldwide human populations (Seielstad *et al.* 1998). Autosomal markers, here represented by Alu insertion polymorphisms, show intermediate values between those found for the mtDNA and the Y chromosome, although Y-chromosome markers exhibit much greater differences between both geographical areas. The Y chromosome behaves as a single locus, and, as such, it is more prone to the vagaries of random drift than a set of independent loci such as the Alu polymorphisms. Moreover, sex-specific population structure (restricted gene flow with isolation) might have enhanced an initial increase of differentiation in male lineages compared to the other of genetic systems analysed.

When focusing in SW Europe, very small (0.1%), non-significant differences were found between Iberians and Italians, in agreement with the large mtDNA homogeneity described in Europe (Simoni *et al.* 2000a, 2000b; Helgason *et al.* 2000; see also Richards *et al.* 2002). Within NW Africa and grouping samples following a geographical east-west criterion (Tunisians and Algerians versus the other populations), no significant differences were found between groups, and 1.3% of the variation was attributed to differences among populations within groups. Nevertheless, when L lineages were removed, the variation attributed to differences among the two groups became significant (1.96%, $p < 0.05$), and the differences within groups did not differ from zero, which points to an east-west differentiation in NW Africa that may have been partially dampened by gene flow from sub-Saharan Africa to both subregions.

Genetic Landscape

The genetic relationship between NW Africans, Iberians and Italians was assessed through a principal co-ordinate analysis based on the distance matrix. The plot of the first two principal co-ordinates (Figure 4a) accounts for 65.8% of the genetic variance observed. The first co-ordinate (56.3%) separates NW African and European populations, except for Moroccan Berbers who are embedded within Europeans, placing Mozabites and Mauritians at one edge and Basques and Galicians at the opposite one. When L lineages are excluded from the analysis (Figure 4b), the plot clusters Europeans, Moroccan Berbers and Southern Berbers and even Tunisians in a group, whereas Algerians, Mozabites and Mauritians are more distant to this cluster.

The genetic relationships between the populations was also assessed through a principal component analysis based on the frequencies of the haplogroups displayed (Richards *et al.* 2002). We observed a similar general pattern to that displayed in the principal co-ordinates, but some differences were also found. The first two principal components (Figure 5a) account for 36.3% of the genetic variance observed and separate the NW African populations, characterised by high frequencies of L and U6 sequences (with absolute correlations of 0.619 for L1, 0.887 for L2, 0.781 for L3, and 0.663 for U6), from the rest of populations, which present high frequencies of H lineages (with an absolute correlation of 0.835). The second principal component encompassed 12.1% of the genetic variance observed and separated the Southern Italians, Tuscans and Sicilians from the rest of the SW Europeans by their low frequencies of K (absolute correlation of 0.458) and the presence of J/T lineages (absolute correlation of 0.735) in their genetic pool. When L sequences were removed from the analysis (Figure 5b), the first two principal components encompassed 33.4% of the genetic variance and separated most NW African populations from Italians, with the remaining populations lying between them.

Discussion

The phylogeographic analysis of mtDNA in the Western Mediterranean has shown the presence of a common set of haplogroups shared with the rest of Europe and

the Middle East (H, J, T, U, I, W, X), plus those of probable local origin (U6, V), and others introduced by gene flow from the south (L) and east (M). In this respect, our regional study, which has gathered published and new samples, not previously jointly analysed, confirms the basic frame described by Richards *et al.* (2000) for Europe and by Rando *et al.* (1998) for NW Africa. It should be noted, though, that inferring haplogroups from HVRI sequences and three coding-region SNPs could lead to slight imprecisions in the allocation of sequences to haplogroups. For instance, although we have assigned all CRS (Cambridge Reference Sequence) sequences to haplogroup H, 1.5% of all CRS sequences in West Eurasia belong to haplogroup HV* and 3.9% to U* (Richards *et al.* 2000). Typing of SNP 7028 could help in resolving this ambiguity, which nonetheless affects a relatively small number of sequences.

An additional caveat that should be taken into account throughout the discussion is that, although we define our area of study as the Western Mediterranean, for some areas, such as southern France, Corsica, northern Italy and the Kabyle in northern Algeria, no HVRI sequences are available. It is likely that such missing data would refine some of the conclusions we will reach below.

Now, we will discuss in detail the phylogeographic pattern for NW Africa, Iberia and Italy, and the trans-mediterranean gene flow.

Northwest African mtDNA Landscape

The main difference, found through the mtDNA analysis, between the populations of the two geographical areas studied is the presence of sub-Saharan L lineages in NW Africa compared to SW Europe, up to the point that, if L sequences were removed from the analyses, most NW African populations were genetically very close to SW Europeans. Since L sequences make up almost all mtDNA lineages in sub-Saharan Africa, and particularly in the areas just to the south of NW Africa, the frequency of L haplogroups in NW Africa can be read directly as a measure of gene flow. Thus, it can be estimated that $25.9 \pm 2.1\%$ of the NW African mtDNA pool has a sub-Saharan origin, under the assumption of negligible back flow from NW to sub-Saharan Africa. A similar estimation can be

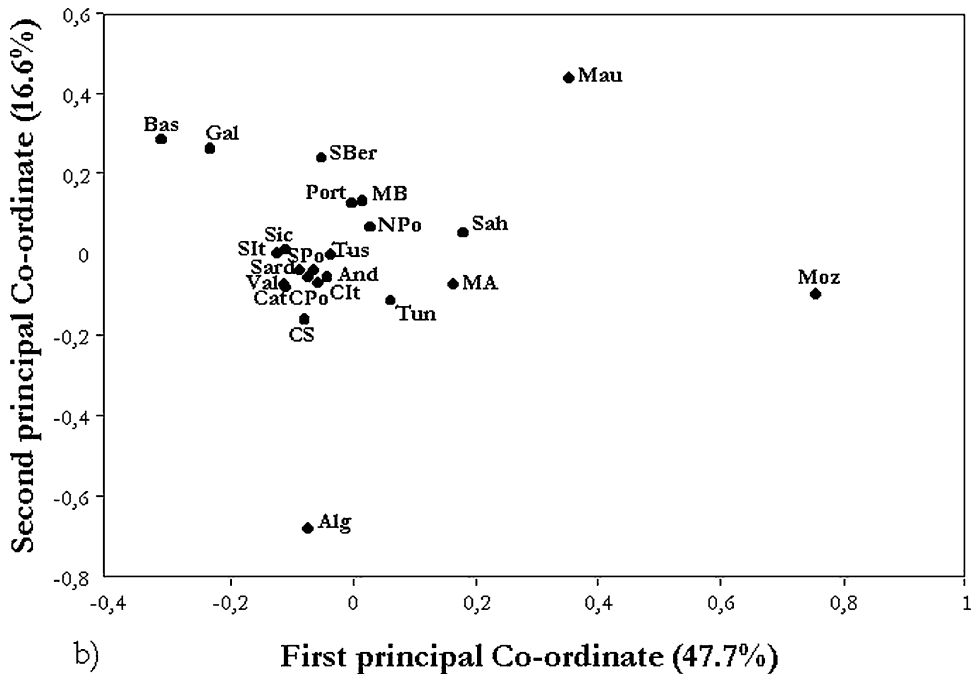
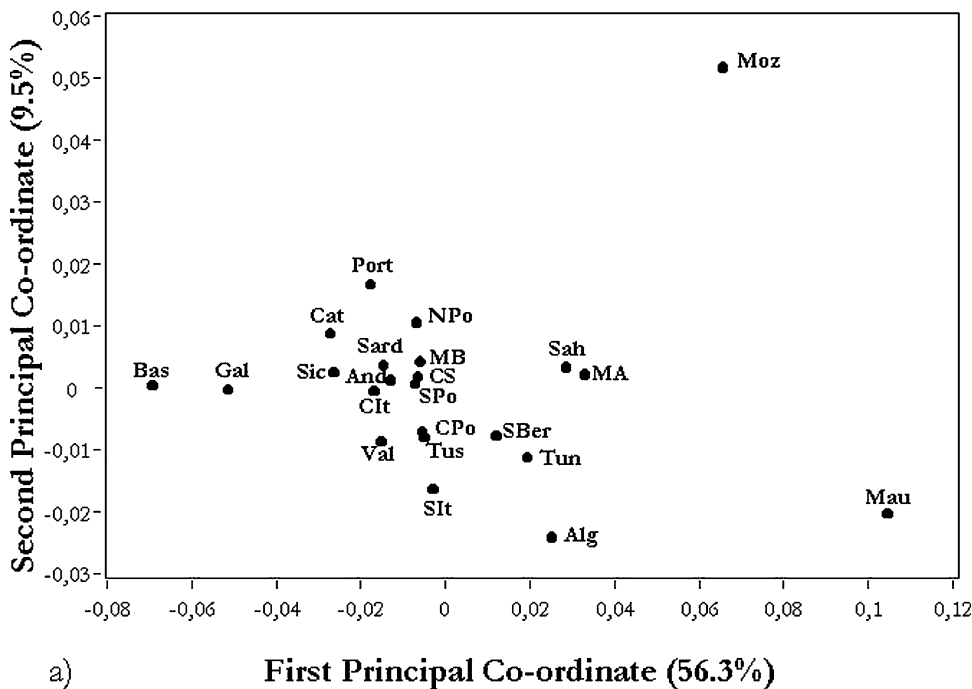
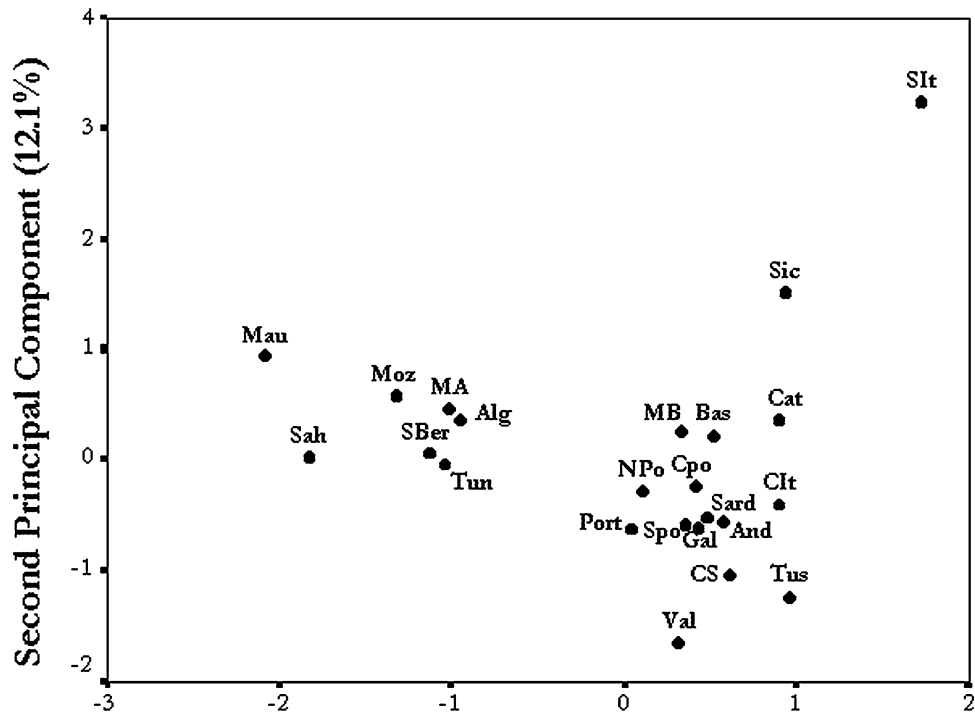
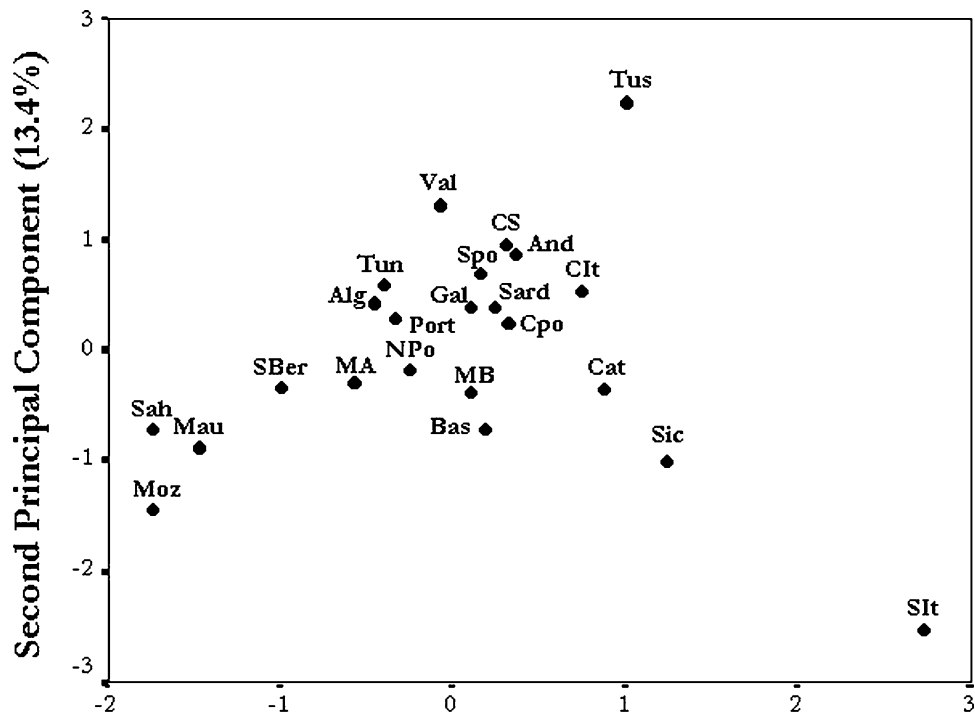


Figure 4 Plot of the two principal co-ordinate (PC) scores based on the genetic distance matrix of NW African, Iberian and Italian samples. a) PC analysis of populations considering all sequences, and b) PC analysis of populations without sequences belonging to haplogroups L. Abbreviations: Alg: Algerians; And: Andalusians; Bas: Basques; Cat: Catalans; CIt: Central Italians; Cpo: Central Portuguese; CS: Central Spaniards; Gal: Galicians; MA: Moroccan Arabs; Mau: Mauritians; MB: Moroccan Berbers; Moz: Mozabites; NPo: Northern Portuguese; Port: Portuguese; Sah: Saharawis; Sard: Sardinians; SBer: Southern Berbers; Sic: Sicilians; SIt: Southern Italians; Spo: Southern Portuguese; Tun: Tunisians; Tus: Tuscans; Val: Valencians.



a) First Principal Component (24.2%)



b) First Principal Component (20.0%)

Figure 5 Plot of the two first Principal component (PC) scores based on the haplogroup frequencies of NW African, Iberian and Italian samples. a) PC analysis of populations considering all sequences, and b) PC analysis of populations without sequences belonging to L haplogroups. Abbreviations as in Fig. 4.

performed for Y-chromosome lineages, since E1* and E3a* haplogroups (according to the nomenclature of the Y Chromosome Consortium, 2002) found in NW Africa at a frequency of $8.0\% \pm 2.0\%$ (Bosch *et al.* 2001), are of sub-Saharan origin. The female- and male-mediated estimates of sub-Saharan gene flow into NW Africa are clearly different, which could be a local consequence of a global trend to higher female than male migration (Salem *et al.* 1996; Seielstad *et al.* 1998; Pérez-Lezaun *et al.* 1999). Autosomal markers such as Alu insertion polymorphisms also show frequency patterns compatible with gene flow from sub-Saharan Africa into NW Africa (Comas *et al.* 2000), although the absence of a clear phylogeographic structure in that case prevents the estimation of gene flow without specifying a parental, non-admixed population for NW Africa.

Within NW Africa, L sequences are most frequent in Mauritians and Saharawi, whereas their frequency is lowest in northern populations. Alu insertion polymorphism analysis in NW Africa (Comas *et al.* 2000) has also shown that gene flow from sub-Saharan Africa in the southern part of this geographical area was more pronounced. A similar genetic gradient was also observed in NE Africa along the Nile valley from analysing Egyptian and Nubian mtDNA sequences (Kriings *et al.* 1999), where south-north migration (and vice versa) could be facilitated by the Nile.

Sequence frequency and diversity, and nucleotide diversity, point to NW Africa as the cradle of U6, with an estimated age of $47,000 \pm 18,000$ years. Such an ancient age contrasts with the limited spread of U6, which is found in N Africa, the Canaries and Iberia, and at very low frequencies in Italy, the Middle East, and the Sahel. This could be explained because, with the exception of the Moslem invasions of Iberia and Sicily, no large population expansion has been known to originate in NW Africa, and the gene tree structure for U6 does not seem compatible with a strong population expansion. U6 represents, thus, a local background in NW Africa. Its relatively low frequency ($\sim 10\%$ overall, although ranging from absence in Algeria to 28.2% in the Mozabites) is in stark contrast with the high frequency of Y-chromosome haplogroup E3b2* (64%; Bosch *et al.* 2001), which may also have originated (or expanded to such high frequency) locally in NW Africa. This discrepancy may be the result of ancient, random,

locus-specific drift, and/or of a male-biased bottleneck or migration. A locus-specific effect may be evidenced by the fact that AMOVA between Iberian and NW African populations is much higher for Y chromosome haplogroups than for multiple autosomal Alu insertion polymorphisms or mtDNA. Since men contribute their autosomes as well, the fact that population differentiation as demonstrated by autosomal loci is much closer to that for mtDNA than to that for the Y chromosome may be taken as evidence for ancient, random, locus-specific drift affecting the Y chromosome.

NW African populations are relatively heterogeneous in their mtDNA sequence pools. The eastern populations (Algeria and Tunisia) may have received more gene flow from the east, as evidenced by the frequencies of M1. This haplogroup originated in East Africa (Quintana-Murci *et al.* 1999) with a frequency $\sim 20\%$ in Ethiopians (Passarino *et al.* 1998), and declines north-westwards (Nubians $\sim 10\%$ and Egyptians $\sim 8\%$; Kriings *et al.* 1999), whereas its frequency in the Middle East is lower ($\sim 3\%$ in Jordanians from Amman, Richards *et al.* 2000; $\sim 2\%$ Israeli Palestinians, Richards *et al.* 2000; $\sim 2\%$ in Israeli Druze, Macaulay *et al.* 1999).

The major outlier within NW Africa are the Mozabites, a well-known Berber isolated group in Algeria, where drift may have altered haplogroup frequencies.

SW European mtDNA Landscape

The mtDNA homogeneity observed in Europe (Simoni *et al.* 2000a and 2000b; Helgason *et al.* 2000, see also Richards *et al.* 2002) is also seen in the present analysis of the West Mediterranean samples, and contrasts with the heterogeneity of NW African populations. All the European samples present the same set of haplotypes with similar frequencies, short genetic distances to each other, and no clear genetic structure, up to the point that populations from Iberia and Italy do not each form a neat group. It should be noted that this homogeneity is seen at the current level of phylogenetic resolution, and that a more fine-grained structure may emerge from the analysis of complete mtDNA sequences (Richards *et al.* 2002).

The most outstanding feature in the west Mediterranean genetic landscape is the outlier position of Sardinians and Basques shown by classical genetic markers

(Cavalli-Sforza *et al.* 1994; Calafell & Bertranpetit 1994; Cappello *et al.* 1996) and Y-chromosome polymorphisms (Caglià *et al.* 1997; Scozzari *et al.* 2001; Bosch *et al.* 2001), although not so pronounced in the Basques. Nevertheless, mtDNA data reveals no differences between these two populations and the rest of European populations. This has also been shown in Basques by analysis of 11 Alu insertion polymorphisms in west Mediterranean populations (Comas *et al.* 2000).

Genetic Exchange Through the Mediterranean

Each of the subregions analysed (NW Africa and SW Europe) shows sequences that originated on the opposite shore of the Mediterranean. This is particularly clear in the case of U6 and L in SW Europe. L sequences are found at frequencies $\sim 3\%$ in Iberia and $\sim 2.4\%$ in Italy. Given the relatively high frequencies of L sequences in NW Africa, it is not clear whether they were contributed by the historical populations movements from the south to the north of the Mediterranean (such as the Moslem invasions of the 7th–11th centuries), or whether its presence is associated with other processes not directly linked to NW Africa. Out of 23 different L sequences in Iberia, two were also found in NW Africa (as well as in sub-Saharan Africa), and 7 others were found in sub-Saharan Africa (in a dataset comprising 1,158 individuals from 20 populations; Graven *et al.* 1995, Pinto *et al.* 1996; Watson *et al.* 1996; Mateu *et al.* 1997; Rando *et al.* 1998; Krings *et al.* 1999; Pereira *et al.* 2001; Brehm *et al.* 2002) but not in NW Africa. Treating the set of L sequences in Iberia as if it were a population reveals genetic distances from some W African populations, such as the Senegalese and Yoruba, that are slightly smaller than those between L sequences in Iberia and NW Africa. Thus, it may be the case that gene flow from NW Africa is not entirely responsible for the presence of L sequences in Iberia.

This may be even clearer in Italy, where the frequency of U6 is much lower than in Iberia (one out of 411 individuals), and where none of the eight L sequences has been found in NW Africa. Three Italian L sequences have been described throughout Africa, and the remaining five are not found in $>1,000$ sub-Saharan individuals. Thus, the presence of L sequences cannot

be attributed to migration from NW Africa, and may instead represent gene flow from other sources, such as the Neolithic expansion or the Roman slave trade.

In contrast to mtDNA, no sub-Saharan Y chromosomal lineages were detected in Iberia (Bosch *et al.* 2001), or in Italy (Rosser *et al.* 2000), although sample sizes in these studies (97 and 99 chromosomes respectively) may not be sufficient to rule out their presence at low frequencies.

As hinted above, the presence of haplogroup U6 in Iberia may signal gene flow from NW Africa, and those of the subhaplogroup U6b1 recent gene flow from the Canary Islands. Haplogroup U6 is present at frequencies ranging from 0 to 7% in the various Iberian populations, with an average of 1.8%. Given that the frequency of U6 in NW Africa is $\sim 10\%$, the mtDNA contribution of NW Africa to Iberia can be estimated at 18%, with a 95% confidence interval of 8%–26% (estimated by sampling with replacement 10,000 times in populations having the same sample sizes and U6 frequencies as Iberia and NW Africa). This is larger than the contribution estimated with Y-chromosomal lineages (7%, 95% confidence interval 1%–14%, Bosch *et al.* 2001). However, it should be noted that the variance due to genetic drift is not included in the estimates, and this may have had a larger effect on U6, which has a much lower frequency in NW Africa than its Y-chromosome counterpart, E3b2*. In the same way, we can estimate the Canarian female contribution to the Iberian Peninsula: the subhaplogroup U6b1 is present at a frequency of 13% in the Canary Islands, and reached a frequency of 0.2% in the Iberian Peninsula. Thus, the mtDNA lineages of the Canary Islands contributed 1.5%, with a 95% confidence interval 0–4.7%, to the genetic pool of Iberia. The presence of lineages belonging to the U6b1 haplogroup in the Iberian Peninsula suggests recent gene flow from the Canary Islands, due to recent migration or to the enslavement and deportation of the native Canarians (also called Guanches) at the time of conquest by the kingdom of Castile (15th century).

With the present data, and in conjunction with other loci, we have glimpsed the palimpsest history of the Western Mediterranean; in that history, the geographical barriers imposed by the Sahara Desert and the Mediterranean Sea might not have been strong enough to prevent a certain degree of gene flow among already

differentiated populations, as they were not barriers to the flow of cultures, languages, and religions.

Acknowledgments

We express our appreciation to the blood donors who participated in the present study. We especially thank Antonio Alonso (Instituto Toxicológico, Madrid), Manuel Crespillo (Instituto Nacional de Toxicología, Barcelona), Manuel López-Soto (Instituto Nacional de Toxicología, Sevilla), Olga Rickards (Università Tor Vergata, Rome), and Martin Richards (Huddersfield University) for providing unpublished data. This work was supported by Dirección General de Investigación, Ministerio de Ciencia y Tecnología in Spain grant PB98-1064 and BOS2001-0794, and by Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya grant 2001SGR00285 and fellowship 2000FI00696 to SP.

References

- 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.
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147.
- Arnaiz-Villena, A., Benmamar, D., Alvarez, M., Diaz-Campos, N., Varela, P., Gomez-Casado, E. & Martinez-Laso, J. (1995) HLA allele and haplotype frequencies in Algerians. Relatedness to Spaniards and Basques. *Hum Immunol* **43**, 259–268.
- Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Prasad, B.V., Reddy, P.G., Rasanayagam, A., Papiha, S.S., Villems, R., Redd, A.J., Hammer, M.F., Nguyen, S.V., Carroll, M.L., Batzer, M.A. & Jorde, L.B. (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* **11**, 994–1004.
- 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., 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., Perez-Lezaun, A., Comas, D., Mateu, E. & Bertranpetit, J. (1997) Population history of north Africa: evidence from classical genetic markers. *Hum Biol* **69**, 295–311.
- Bosch, E., Calafell, F., Perez-Lezaun, A., Clarimon, J., Comas, D., Mateu, E., Martinez-Arias, R., Morera, B., Brakez, Z., Akhayat, O., Sefiani, A., Hariti, G., Cambon-Thomsen, A. & Bertranpetit, J. (2000) Genetic structure of north-west Africa revealed by STR analysis. *Eur J Hum Genet* **8**, 360–366.
- 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.
- Brakez, Z., Bosch, E., Izaabel, H., Akhayat, O., Comas, D., Bertranpetit, J. & Calafell, F. (2001) Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* **28**, 295–307.
- Brehm, A., Pereira, L., Bandelt, H.J., Prata, M.J. & Amorim, A. (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann Hum Genet* **66**, 49–60.
- Cabot *et al.* (1988) *ESEE: The Eyeball Sequence Editor, version 1.06*. Burnaby: University of British Columbia.
- Calafell, F. & Bertranpetit, J. (1994) Principal component analysis of gene frequencies and the origin of Basques. *Am J Phys Anthropol* **93**, 201–215.
- Cali, F., Le Roux, M.G., Anna, R., Flugy, A., De Leo, G., Chiavetta, V., Ayala, G.F. & Romano, V. (2001) MtDNA control region and RFLP data for Sicily and France. *Int J Legal Med* **114**, 229–231.
- Caglià, A., Novelletto, A., Dobosz, M., Malaspina, P., Ciminelli, B.M. & Pascali, V.L. (1997) Y-chromosome STR loci in Sardinia and continental Italy reveal islander-specific haplotypes. *Eur J Hum Genet* **5**, 288–292.
- Cappello, N., Rendine, S., Griffio, R., Marnelli, G.E., Succa, V., Vona, G. & Piazza, A. (1996) Genetic analysis of Sardinia: I. Data on 12 polymorphisms in 21 linguistic domains. *Am Hum Genet* **60**, 125–141.
- Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1994) *History and geography of human genes*. Princeton, NJ: Princeton University Press.
- Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A. & Bertranpetit, J. (1998) HLA evidence for the lack of genetic heterogeneity in Basques. *Ann Hum Genet* **62**, 123–132.
- Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefranc, G., Stoneking, M., Batzer, M.A., Bertranpetit, J. & Sajantila, A. (2000) Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* **107**, 312–319.
- Côrte-Real, H.B., Macaulay, V.A., Richards, M.B., Hariti, G., Issad, M.S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B.C. (1996) Genetic diversity in

- the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331–350.
- Crespillo, M., Luque, J.A., Paredes, M., Fernandez, R., Ramirez, E. & Valverde, J.L. (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. *Int J Legal Med* **114**, 130–132.
- Di Rienzo, A. & Wilson, A.C. (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* **88**, 1597–1601.
- Efron, B. (1982) *The Jackknife, the bootstrap and other resampling plans*. Philadelphia, Pa: Society for industrial and Applied Mathematics.
- Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.
- 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.
- Gower, J.C. (1966) Some distance properties of latent root and vector methods in multivariate analysis. *Biometrika* **27**, 857–874.
- Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. & Excoffier, L. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* **12**, 334–345.
- Gresham, D., Mora, B., Underhill, P.A., Passarino, G., Lin, A.A., Wise, C., Angelicheva, D., Calafell, F., Oefner, P.J., Shen, P., Tournev, I., Pablo, R., Kucinskis, V., Perez-Lezaun, A., Marushiakova, E., Popov, V. & Kalaydjieva, L. (2001) Origins and divergence of the Roma (gypsies). *Am J Hum Genet* **69**, 1314–1331.
- Helgason, A., Sigurðardóttir, S., Gulcher, J.R., Ward, R. & Stefansson, K. (2000) MtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* **66**, 999–1016.
- Krings, M., Salem, A.E., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., Sajantila, A., Paabo, S. & Stoneking, M. (1999) MtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet* **64**, 1166–1176.
- Liegeois, J.P. (1994) *Roma, Gypsies, Travellers*. Strasbourg, France: Council of Europe Press.
- López-Soto, M. & Sanz, P. (2000) Mitochondrial DNA polymorphisms in individuals living in Andalusia (south of Spain) and Extremadura (western Spain). In: *Cuadernos de Medicina Forense*, pp. 17–24.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-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.
- Mateu, E., Comas, D., Calafell, F., Perez-Lezaun, A., Abade, A. & Bertranpetit, J. (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. *Ann Hum Genet* **61**, 507–518.
- Morral, N., Bertranpetit, J., Estivill, X., Nunes, V., Casals, T., Gimenez, J., Reis, A., Varon-Mateeva, R., Macek, M. Jr. & Kalaydjieva, L., et al. (1994). The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat Genet* **7**, 169–175.
- Nei, M. (1987) *Molecular evolutionary genetics*. New York: Columbia University Press.
- Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M. & Santachiara-Benerecetti, A.S. (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* **62**, 420–434.
- 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.
- Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J. & Amorim, A. (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* **65**, 439–458.
- Perez-Lezaun, A., Calafell, F., Comas, D., Mateu, E., Bosch, E., Martinez-Arias, R., Clarimon, J., Fiori, G., Luiselli, D., Facchini, F., Pettener, D. & Bertranpetit, J. (1999) Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am J Hum Genet* **65**, 208–219.
- Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga J.M. & Cabrera, V.M. (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* **60**, 321–330.
- Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A.S. (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* **23**, 437–441.
- Rando, J.C., Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga, J.M., Cabrera, V.M. & Bandelt, H.J. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* **62**, 531–550.
- Rando, J.C., Cabrera, V.M., Larruga, J.M., Hernandez, M., Gonzalez, A.M., Pinto, F. & Bandelt, H.J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* **63**, 413–428.
- 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. (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.
- Rickards, O., Martínez labarga, C., Casalotti, R., Castellana, G., Tunzi sisto, A.M. & Mallegni, F. (2000) MtDNA variability in extinct and extant populations of Sicily and southern Italy. In: C. Renfrew & K. Boyle (Eds), *Archaeogenetics: DNA and the population prehistory of Europe*, pp. 175–183, McDonald Institute Monographs, McDonald Institute for Archaeological Research, Cambridge.
- Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D.G., Brede, G., Cooper, G., Cortereal, 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., Sallard, 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.
- Saillard, J., Magalhaes, P.J., Schwartz, M., Rosenberg, T. & Norby, S. (2000) Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. *Hum Biol* **72**, 1065–1068.
- 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.
- Salem, A.H., Badr, F.M., Gaballah, M.F. & Paabo, S. (1996) The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am J Hum Genet* **59**, 741–743.
- 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.
- Schneider, S., Kueffer, J.M., Roessli, D. & Excoffier, L. (1996) *Arlequin (ver1.0): a software environment for the analysis of population genetics and Biometry Lab*. Geneva, Switzerland: University of Geneva.
- Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Latini, V., Varesi, L., Memmi, M.M., Romano, V., De Leo, G., Gennarelli, M., Jaruzelska, J., Villems, R., Parik, J., Macaulay, V. & Torroni, A. (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* **62**, 871–884.
- Simoni, L., Gueresi, P., Pettener, D. & Barbujani, G. (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. *Hum Biol* **71**, 399–415.
- Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. & Barbujani, G. (2000a) Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet* **66**, 262–278.
- Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. & Barbujani, G. (2000b) Reconstruction of prehistory on the basis of genetic data. *Am J Hum Genet* **66**, 1177–1179.
- 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., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L. & Wallace, D.C. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.
- Torroni, A., Bandelt, H.J., 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**, 1137–1152.
- Torroni, A., Bandelt, H.J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martínez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M.L., Huoponen, K., Laitinen, V., Koivumaki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A.S., Semino, O. & Scozzari, R. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69**, 844–852.
- Watson, E., Bauer, K., Aman, R., Weiss, G., Haeseler, A. & Paabo, S. (1996) MtDNA sequence diversity in Africa. *Am J Hum Genet* **59**, 437–444.
- Watson, E., Forster, P., Richards, M. & Bandelt, H.J. (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* **61**, 691–704.
- Y Chromosome Consortium, The (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* **12**, 339–348.

Received: 10 October 2002

Accepted: 19 March 2003