Coll. Antropol. **34** (2010) 4: 1215–1228
Original scientific paper

# Searching the Peopling of the Iberian Peninsula from the Perspective of Two Andalusian Subpopulations: A Study Based on Y-chromosome Haplogroups J and E

**Beatriz Ambrosio[1], Candela Hernández[1], Andrea Novelletto[3], Jean M. Dugoujon[4], Juan N. Rodríguez[5], Pedro Cuesta[2], César Fortes-Lima[1] and Rosario Calderón[1]**

[1] Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense, Madrid, Spain
[2] Centro de Proceso de Datos, Unidad de Apoyo a la Investigación, Universidad Complutense, Madrid, Spain
[3] Dipartimento di Biologia, Università »Tor Vergata« Rome, Italy
[4] Laboratoire d'Anthropologie, FRE 2960, Centre National de la Recherche Scientique (CNRS), Université Paul Sabatier, Toulouse, France
[5] Servicio de Hematología, Hospital Juan Ramón Jiménez, Huelva, Spain

**A B S T R A C T**

*This study aims at a high-resolution analysis of Y-chromosome J and E haplogroups among Andalusians to reconstruct Neolithic, protohistorical and historical migrations in the Mediterranean region. Genotyping of two samples from Granada (n=250 males) and Huelva (n=167 males) (Spain) with Y-chromosome binary and microsatellite markers was performed, and the results compared with other Mediterranean populations. The two samples showed genetic differences that can be associated with different evolutionary processes. Migrations toward Andalusia probably originated in the Arabian Peninsula, Fertile Crescent, Balkan region and North Africa, and they would have predominantly occurred in protohistoric and historic times. Maritime travel would have notably contributed to recent gene flow into Iberia. This survey highlight the complexity of the Mediterranean migration processes and demonstrate the impact of the different population sources on the genetic composition of the Spanish population. The main in-migrations to Iberia most likely did not occur through intermediate stages or, if such stages did occur, they would have been very few.*

***Key words**: human genetic history, population structure, migrations, Andalusia, Mediterranean region*

## Introduction

The strategic position of the Iberian Peninsula, which covers 582,860 km$^2$ and is situated at the most southwestern end of Europe, its proximity to the African continent, and its orographic characteristics, which make movements within the Peninsula difficult, have had striking implications for the successive processes of human settlement throughout prehistory and history. During the Last Glacial Maximum (LGM), the Iberian Peninsula was one of the few European refuges for the human population, and after the improvement of climate conditions, population dispersals occurred towards other European regions. Subsequently, in prehistoric and historic times, several population movements took place from the Eastern Mediterranean, North Africa and Europe to the Peninsula. These inputs were particularly intense in the most southern region of the Peninsula (Andalusia region) and in the Spanish Levante (Valencia region and its adjacent coasts) and their genetic effects could have been strongly affected by the considerable population size of the Peninsula, the limited population size of northwest African descent, and the demography of the easternmost Mediterranean from prehistory through the largest part of the Antiquity. Consequently, the ethnic profile of the Iberian Peninsula should be considered rich and complex. Archaeological evidence suggests that around 4,000 BC, Iberia was most densely populated along its Mediter-

Received for publication October 1, 2010

ranean and South Atlantic coasts (stretching for 3,080 km). Population density in these regions has been estimated[1] to be ~2–5 inhab/km$^2$.

Two Andalusian subpopulations originating of the Huelva and Granada provinces were selected to carry out the present survey. Huelva is the most southwestern Andalusian province, and Granada is geographically located just on the eastern side of the region (see Figure 1). Both territories are mountainous in part and are located along the Andalusian coastal fringe. The territories are approximately 300 km apart and are more than 1,000 km away from the Pyrenees. Furthermore, Huelva and Granada have a long, important and differentiated history. Huelva is identified with the ancient, protohistorical Tartessian civilisation. The city and kingdom of Tartessos (ca. 800–540 BC) were presumably situated in the Huelva region[2]. Since the 13th century, Huelva shared a border with Portugal. Granada is especially known for the Nazari Kingdom (13th–15th centuries), which extended the Muslim domination in that region for two and a half centuries longer than in the rest of the Peninsula, and whose cultural and ethnic repercussions were especially profound[3]. During the first half of the Middle Ages, the high population density of Andalusia contrasted significantly with the semi-depopulated lands of the north of the Peninsula, which had no major cities. Thorough information on the prehistory, history, territory, and geography of Huelva can be found in an earlier survey[4].

The significant number of studies performed in the last five years on the phylogeography of the major clades of the human Y chromosome tree [haplogroups (Hgs)], based mostly on single nucleotide polymorphisms (Y-SNPs) and some insertion/deletion polymorphisms (small indels), has shown that, in general, geography seems to be the main factor that has shaped the genetic diversity patterns observed within and among continents[5]. The distribution of specific haplogroups is restricted to defined continents or major areas, and hence contributes to the structuring of large contemporary human populations[6–9]. Furthermore, the continuing discovery of Y-SNPs that define new lineages internal to these larger clades are leading to a more refined and complex phylogeny, which allows the detection of subtle signatures of recent admixtures as a result of demographic events, and the unveiling of genetic distinctions among geographical sub-regions.

By virtue of extensive studies Hgs E and J share the above properties, allowing researchers to evaluate the degree of phylogenetic diversification also on small spatial scales (e.g. Mediterranean sub-regions). Some of the lineages or sub-lineages of E and J have reached the Mediterranean, occurring in informative frequencies among their populations, especially in populations of coastal areas. These results are currently being interpreted in terms of directionality, intensity and antiquity of the migratory processes and their relevance in the continuous genetic flow inside this geographic space[10–14].

The available data on Y-chromosome binary haplogroup diversity in Spanish populations are relatively abundant, even though they are far from being homogeneous with respect to the levels of genealogical resolution reached, so that it would make necessary calling for their refinement. The diversity of paternal lineages in the Andalusian population samples indicates that these populations have integrated multiple migrations and that the sources of gene flow appear to be more intense and diverse in the west than in the east of Andalusia. The results recently published on the E haplogroup for the autochthonous population of Huelva[15] have revealed high genetic diversification of this major clade, with a frequency of 3% of the E-M81 subclade, which is commonly referred to as the »Berber marker«. The appreciable representation (4%) of the E-V13 lineage, which has maximum frequency in the Balkans, and the co-occurrence of E-M34 (1.34%), which is clearly prevalent among Jews, constitute other interesting results.

The present study aimed at performing a high-resolution analysis of Y-chromosome J and E haplogroups and their internal diversities among Andalusians from Granada province, and on the J haplogroup in the population sample from Huelva province to discern in which way those lineages were introduced in the Iberia Peninsula through Andalusia. The genetic data we produced were further analysed in terms of population sub-structuring; our results are framed within the complex demographic and historical context of the region.

## Materials and Methods

### *Populations, sampling process and population samples*

The map of Andalusia, the location of the two studied provinces and the municipalities where the sampling was performed are displayed in Figure 1. Sampling localities were selected following regional and demographic stability criteria. Both capital cities and the municipalities situated along the coast were avoided because of their intense demographic growth or their tourist relevance. The number of healthy, unrelated autochthonous males sampled and genotyped in the current study was 250 from Granada and 167 from Huelva. The collection of samples took place between 2004 and 2008. Blood samples were taken by medical staff associated with the Juan Ramón Jiménez Hospital in Huelva, and the Blood Transfusion Centre in Granada. In each phase of our fieldwork (11 in total), two members of our research group (RC and BA) participated in the sample collection process. Each donor was carefully informed about the goals of this research project. After procuring the appropriate informed consent in accordance with the Spanish Legislation on Biomedical Research, blood samples were taken from subjects representing at least the third generation born in the same province.

Genomic DNA was isolated from approximately 5–7 mL of fresh, whole, EDTA-treated blood using a standard proteinase-K digestion followed by phenol-chloroform extraction and ethanol precipitation.

*Genotyping of haplogroups*

For the analysis of the E and J Y-DNA haplogroups and their subclades, we have followed the Y Chromosome Consortium (YCC) nomenclature[7] (http://ycc.biosci.arizona.edu/). Updated information on Y-DNA chromosome haplogroups and their subclades and nomenclature can be found at http://www.isogg.org/.

The Granada population sample was first analysed for the presence/absence of the Y-chromosome Alu polymorphism (YAP element, DYS287)[16]. All YAP (+) individuals were tested to search for the presence of SNPs internal to Hg E, following the preceptive hierarchical order. The remaining YAP (–) Granada samples, together with those from Huelva, were further genotyped for the 12f2 marker[17] to identify haplogroup J. The Huelva sample had been characterised earlier for haplogroup E[15]. To investigate Hg E in the Granada sample, protocols identical to those used in the Huelva sample were applied. The total number of binary polymorphisms (SNPs) used to characterise Hg E was sixteen.

For the analysis of Hg J a total of nine SNPs, M267, M172, M365, M367, M369, M410, M67, M12 and M241, were genotyped using either RFLP detections or direct sequencing following a hierarchical order. All of these SNPs were amplified using previously published primers[11,18]. The dinucleotide DYS413 and tetranucleotide DYS445 microsatellite loci were analyzed in all M410 derived samples[19] using primers published in[20,21].

*Microsatellite markers*

All (n=417) Andalusian samples were also typed for 16 Y-chromosome specific microsatellites using the AmpFlSTR® Yfiler™ Kit (Applied Biosystems). Alleles of microsatellites or STRs (short tandem repeats) were designated based on the number of variable repeats included[22]. For the Y-STR population comparison analysis, we considered either the »nine loci extended haplotype« (http://www.yhrd.org) without the DYS385a/b microsatellites or the »seven loci minimal haplotype«. The number of repeats in DYS389II was subtracted from DYS389I. As a result, haplogroup diversity was analysed in terms of the observed biallelic markers (Y-SNPs) and the haplotype diversity due to the variation of microsatellites (Y-STRs) associated with each lineage within a haplogroup.

*Data analysis*

Haplogroup (h) and haplotype (H) genetic diversities, shared haplotypes, and AMOVA (Analysis of Molecular Variance) were calculated using the ARLEQUIN 3.1 software[23]. Repeat variance and mean repeat variance were calculated for Y-STR markers[24], using MICROSAT software: http://hpgl.stanford.edu/projects/microsat/. The $\chi^2$ test and the $F_{ST}$ genetic heterogeneity parameter were used to evaluate random deviations of haplogroup frequencies between populations.

Geographical variation for the J and E sub-haplogroups was analysed by Hierarchical Cluster Analysis (HCA) using the statistical program SPAD (Système Portable Pour l'Analyse de Donnés[25]). HCA is a very powerful multivariate tool because it includes both Principal Component Analysis (PCA) and Factorial Analysis (FA), such that it finds clusters of observations within a data set. HCA was performed based on Euclidean distances and Ward's linkage algorithm. For Hg J, we have included data sets from another 37 population samples taken from the literature (see Table 1). Genetic information was based on population frequencies of the following J lineages: J*, J1-M267, J2*, J2a-M410 and J2b- -M12. In the case of Hg E, we used the population database published in[15], but including our new results, which are the first results concerning Hg E among Granada Andalusians. The sample size for each population sample selected was >20.

For the J1 (M267) and J2a (M410) haplogroups, the phylogenetic relationships among 7 microsatellite haplotypes (Minimal Haplotype without DYS385a/b) were obtained by sequentially performing reduced-median and median-joining procedures[26,27] using the NETWORK 4.5 program (http://www.fluxus-engineering.com/sharenet.htm). To reduce reticulations in the network, microsatellites were weighted proportionally to the inverse of the repeat variance observed in each haplogroup. For this analysis, a set of Mediterranean population samples for which the detailed data were available in the literature were used.

Likewise, contour maps were generated for the J1--M267 and J2a-M410 lineages, and data were obtained from the same database used to construct the HCA. The maps were created using the SURFER v.8 geostatistic program (Golden Software, Inc), and irregularly located data were interpolated (gridded) using the Kriging method[28].

The program BATWING[29,30] was used to obtain dating estimates of haplogroup/sub-haplogroup antiquity based on the diversity of 15 microsatellites (Y-STRs), with each lineage treated independently. It is assumed an unbounded single stepwise mutation model for the microsatellite loci and a coalescent process under an exponential model of population growth from an initially constant-size population. The priors for a (the rate of increase of population size) and b (the time of start of population growth) were relaxed to UNIFORM (0.0, 0.04) and (0.0, 1.0), respectively, in order to explore the signature of population growth which is present in the data. For each Y-STRs the mutation rates published in the Y Chromosome Haplotype Reference Database (http://www.yhrd.org) were used.

## Results

The frequency of Hg J and the sub-haplogroups observed in the two examined samples of autochthonous Andalusians (a), together with the Y-STR haplotypes associated with each of the J lineages (b), are shown in Table 2. Haplogroup J occurs at 9.20% in Granada and 7.20% in Huelva. These findings are in agreement with

**TABLE 1**
THE GEOGRAPHICAL DISTRIBUTION OF HAPLOGROUP J AND SOME OF ITS MAIN INFORMATIVE SUBHAPLOGROUPS OBSERVED IN A SET OF WORLDWIDE HUMAN POPULATIONS

| | | | | Hg J | | Frequency of J subhaplogroups (%) | | | Frequency of J2 subhaplogroups (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Populations | ACRN[1] | Region[2] | n | No. | % | J* | J1-M267 | J2-M172 | J2* | J2a-M410 | J2b-M12 | References |
| 1. Spanish Basques | ESBA | IB | 48 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | Semino et al. 2004 |
| 2. Catalans | ESCN | IB | 28 | 1 | 3.57 | 0.00 | 0.00 | 3.60 | 0.00 | 3.60 | 0.00 | Semino et al. 2004 |
| 3. Andalusians (Huelva) | ESAH | IB | 167 | 12 | 7.19 | 0.00 | 2.40 | 4.79 | 0.00 | 2.99 | 1.80 | Present study |
| 4. Andalusians (Granada) | ESAG | IB | 250 | 23 | 9.20 | 0.40 | 2.80 | 6.00 | 0.40 | 4.80 | 0.80 | Present study |
| 5. Andalusians | ESAN | IB | 93 | 8 | 8.60 | 0.00 | 1.10 | 7.60 | 2.20 | 4.30 | 1.10 | Semino et al. 2004 |
| 6. Netherlands | NL | WEU | 34 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | Semino et al. 2004 |
| 7. Italians (Northern Italy) | ITN | CEU | 126 | 13 | 10.32 | 0.00 | 0.79 | 9.58 | 0.00 | 7.20 | 3.38 | Di Giacomo et al. 2004 |
| 8. Italians (Southern Italy) | ITS | CEU | 595 | 149 | 25.04 | 0.00 | 2.00 | 23.00 | 8.68 | 12.38 | 1.95 | Di Giacomo et al. 2004, Semino et al. 2004 |
| 9. Sicilians (Italy) | ITSY | CEU | 42 | 10 | 23.81 | 0.00 | 7.14 | 16.70 | 11.90 | 4.80 | 0.00 | Semino et al. 2004 |
| 10. Sardinians (Italy) | ITSD | CEU | 144 | 18 | 12.50 | 0.00 | 2.78 | 9.78 | 2.78 | 4.90 | 2.10 | Semino et al. 2004 |
| 11. Czechs | CZ | CEU | 75 | 4 | 5.33 | 0.00 | 0.00 | 5.33 | 0.00 | 0.00 | 5.33 | Battaglia et al. 2008 |
| 12. Hungarians | HU | EEU | 53 | 1 | 1.89 | 0.00 | 0.00 | 1.90 | 0.00 | 1.90 | 0.00 | Battaglia et al. 2008 |
| 13. Poles | PL | EEU | 99 | 2 | 2.02 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | Battaglia et al. 2008 |
| 14. Croats | HR | EEU | 89 | 4 | 4.49 | 0.00 | 0.00 | 4.49 | 0.00 | 1.12 | 3.37 | Battaglia et al. 2008 |
| 15. Bosniacs | BA | EEU | 84 | 12 | 14.29 | 0.00 | 2.40 | 12.00 | 0.00 | 8.40 | 3.60 | Battaglia et al. 2008 |
| 16. Slovenians | SI | EEU | 75 | 3 | 4.00 | 0.00 | 1.33 | 2.60 | 0.00 | 2.60 | 0.00 | Battaglia et al. 2008 |
| 17. Bulgarians | BU | EEU | 39 | 9 | 23.08 | 0.00 | 5.13 | 18.03 | 0.00 | 12.90 | 5.13 | Di Giacomo et al. 2004 |
| 18. Albanians | AL | EEU | 111 | 26 | 23.42 | 0.00 | 3.60 | 19.80 | 0.00 | 5.40 | 14.40 | Semino et al. 2004, Battaglia et al. 2008 |
| 19. Greeks | GR | EEU | 341 | 73 | 21.41 | 0.20 | 1.90 | 19.35 | 1.20 | 12.00 | 6.15 | Di Giacomo et al. 2004, Battaglia et al. 2008 |
| 20. Greeks (Crete) | GRC | EEU | 193 | 75 | 38.86 | 0.00 | 8.29 | 30.61 | 0.00 | 27.50 | 3.11 | King et al. 2008 |
| 21. Belarusians | BY | EEU | 39 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | Di Giacomo et al. 2004 |
| 22. Ukrainians | UA | EEU | 92 | 6 | 6.52 | 0.00 | 0.00 | 6.60 | 0.00 | 3.30 | 3.30 | Battaglia et al. 2008 |
| 23. Romanians | RO | EEU | 130 | 15 | 11.54 | 0.00 | 1.50 | 10.10 | 0.00 | 10.10 | 0.00 | Di Giacomo et al. 2004 |
| 24. Russians | RU | EEU | 223 | 4 | 1.79 | 0.00 | 0.45 | 1.35 | 0.00 | 1.35 | 0.00 | Di Giacomo et al. 2004 |
| 25. Turkish | TR | SWA | 893 | 281 | 31.47 | 0.07 | 7.96 | 23.48 | 10.18 | 11.99 | 1.31 | Cinnioglu et al. 2004, Di Giacomo et al. 2004, Semino et al. 2004 |
| 26. Caucasians | CA | SWA | 150 | 51 | 34.00 | 0.00 | 8.73 | 25.83 | 0.00 | 25.18 | 0.65 | Battaglia et al. 2008, Di Giacomo et al. 2004 |
| 27. Ashkenazim Jewish | ILA | SWA | 82 | 31 | 37.80 | 0.00 | 14.60 | 23.21 | 12.20 | 9.80 | 1.22 | Semino et al. 2004 |
| 28. Sephardim Jewish | ILS | SWA | 42 | 17 | 40.49 | 0.00 | 11.90 | 28.61 | 23.81 | 2.40 | 2.40 | Semino et al. 2004 |
| 29. Syrians | SY | SWA | 50 | 16 | 32.00 | 0.00 | 18.00 | 14.00 | 0.00 | 14.00 | 0.00 | Di Giacomo et al. 2004 |
| 30. Iraqi | IQ | SWA | 156 | 79 | 50.64 | 0.00 | 28.20 | 22.50 | 10.20 | 9.70 | 2.60 | Semino et al. 2004 |
| 31. Iran | IR | SWA | 150 | 52 | 34.67 | 0.00 | 11.33 | 23.30 | 0.00 | 20.00 | 3.30 | Regueiro et al. 2006 |
| 32. Qatar | QA | SWA | 72 | 48 | 66.67 | 0.00 | 58.33 | 8.38 | 0.00 | 5.60 | 2.78 | Cadenas et al. 2008 |
| 33. Yemenites | YE | SWA | 62 | 51 | 82.26 | 0.00 | 72.58 | 9.60 | 0.00 | 9.60 | 0.00 | Cadenas et al. 2008 |
| 34. United Arab Emirates | AE | SWA | 164 | 74 | 45.12 | 0.00 | 34.76 | 10.32 | 0.00 | 9.10 | 1.22 | Cadenas et al. 2008 |
| 35. Tunisians | TN | NAF | 73 | 25 | 34.25 | 0.00 | 30.14 | 4.11 | 1.37 | 0.00 | 2.74 | Semino et al. 2004 |
| 36. Egyptians | EG | NAF | 47 | 11 | 23.40 | 0.00 | 12.77 | 10.63 | 2.13 | 8.50 | 0.00 | Di Giacomo et al. 2004 |
| 37. Sudanese | SD | EAF | 40 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | Semino et al. 2004 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 38. Ethiopians (Amhara) | ETA | EAF | 48 | 17 | 35.42 | 0.00 | 33.33 | 2.10 | 0.00 | 2.10 | 0.00 | Semino et al. 2004 |
| 39. Ethiopians (Oromo) | ETO | EAF | 78 | 3 | 3.85 | 0.00 | 2.56 | 1.28 | 0.00 | 0.00 | 1.28 | Semino et al. 2004 |

[1, 2] Populations were denoted by the first two letters used in country code top-level domains for Internet addresses (http://www.iana.org/cctld/) and by another acronym designating their larger geographical origin.

other observations[14,31] when analysing general Andalusian samples.

The geographic distribution of J sub-haplogroups between the Granada and Huelva subpopulations did not exhibit a significant degree of heterogeneity ($F_{ST}$=0.033, p=0.810); the estimated J internal diversities for these populations were, respectively, 0.858±0.046 and 0.849±0.074. $F_{ST}$ was calculated only on haplogroup J data, to evaluate its internal diversity across populations. Thus, the fixation indexes Hg of J reported here are not comparable with those ones that include all chromosomes within a population sample.

Within the Iberian Peninsula, the average frequency of Hg J is approximately 8% and varies from zero (in the Basque region) to 18% (among southern Portuguese) (see[14,32]). In Portugal, this Y-chromosome marker displays a distribution pattern with increasing frequencies from the north to the south[33]. Across Mediterranean Europe, Hg J also seems to show a clinal distribution from the west to the east[11,34,35]. Peak frequencies of the J haplogroup are concentrated in the Middle East and in the neighbouring southwest Asian populations (30–82%). These figures are lower in northern Africa (Hg J: 20% as an average)[13,14,36,37].

One of the two main subclades that shape the J clade is J1-M267, which ranges from 0 to 8% in Europe. In our Andalusian sample, the overall frequency is 2.60%. All J1 (n=11) chromosomes sampled were further genotyped for M365, M367 and M369 binary markers, which define distinct J1 internal lineages. The presence of the J1* paragroup was detected in all cases as none of these three polymorphisms was identified (see Table 2a). This finding closely agrees with other population studies, which have also found that diversity within the J1 clade is very low as compared with the abundance of the internal lineages of the J2 subclade[38].

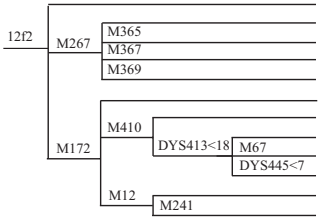
Peak frequencies of J1 Y-chromosomes seem to be restricted[39] to the Arabian region (Qatar: 58%, Yemen: 73%) and the northeast Caucasus (Dagestan) (45–80%)[40]. In Levantine areas, haplogroup J1 reaches frequencies of approximately 31%[41], and in northern Africa, it ranges between 5% and 30%[13,14,36].

Another major lineage within the J clade is J2-M172, which ranges from 0 to 30% in Europe[13,14,35]. In our Andalusian samples, this lineage occurs at frequencies of 5–6%, whereas in other Iberian populations, J2 varies from 3% (in the Basque area) to 15% (in southern Portugal)[32]. In Sicily, J2 is one of the most represented Y-chromosome J sub-haplogroups, being found at frequencies of 15% on the eastern side and 7% on the western side of the island[42]. In northern Africa, this sub-haplogroup is better represented in Egypt (northern Egypt: 9.30%) than in Morocco (2%)[10], and is nearly absent in eastern Asia[43].

Andalusian Y-chromosomes carrying the M172 mutation (n=18) harboured a number of J2 derived lineages

**TABLE 2**
FREQUENCY AND EXTENT OF DIFFERENTIATION OF Y-CHROMOSOME HAPLOGROUP J OBSERVED BOTH WITH BIALLELIC MARKERS (Y-SNPs) (A) AND MICROSATELLITES (Y-STRs) (B) IN THE ANDALUSIAN SUBPOPULATIONS FROM GRANADA AND HUELVA

a)




| Lineages [1] | *Andalusian Subpopulations* *Granada* (n=250) | *Huelva* (n=167) |
|---|---|---|
| J* | 1 (0.40) [3] | 0 |
| J1* | 7 (2.80) | 4 (2.40) |
| J1b | 0 | 0 |
| J1e1 (J1c3a) [2] | 0 | 0 |
| J1e2 (J1c3b) | 0 | 0 |
| J2* | 1 (0.40) | 0 |
| J2a* | 1 (0.40) | 0 |
| J2a1* (J2a4*) | 4 (1.60) | 1 (0.60) |
| J2a1b (J2a4b) | 4 (1.60) | 1 (0.60) |
| J2a1h (J2a4h1a) | 3 (1.20) | 3 (1.80) |
| J2b* | 1 (0.40) | 1 (0.60) |
| J2b2 | 1 (0.40) | 2 (1.20) |
| | 23 (9.20) | 12 (7.20) |

[1] Karafet et al. 2008
[2] Nomenclature in parentheses: ISOGG 2010
[3] Relative frequencies

b)

| | Haplotype | Haplogroup | DYS19 | DYS389 I | DYS389 II | DYS390 | DYS391 | DYS392 | DYS393 | DYS385a/b | DYS438 | DYS439 | *N* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Granada* | H1 | J*-12f2 | 15 | 12 | 28 | 24 | 10 | 11 | 12 | 15,18 | 9 | 12 | ***1*** |
| | H2 | J1*-M267 | 14 | 12 | 29 | 23 | 10 | 11 | 12 | 13,18 | 10 | 11 | ***1*** |
| | H3 | J1*-M267 | 14 | 13 | 26 | 23 | 10 | 11 | 12 | 14,17 | 10 | 12 | ***2*** |
| | H4 | J1*-M267 | 14 | 13 | 29 | 23 | 10 | 11 | 12 | 14,18 | 10 | 12 | ***1*** |
| | H5 | J1*-M267 | 14 | 13 | 31 | 22 | 10 | 11 | 12 | 14,18 | 10 | 12 | ***1*** |
| | H6 | J1*-M267 | 14 | 14 | 30 | 23 | 10 | 11 | 12 | 13,20 | 10 | 11 | ***2*** |
| | H7 | J2*-M172 | 14 | 13 | 31 | 22 | 10 | 11 | 12 | 14,18 | 10 | 12 | ***1*** |
| | H8 | J2a*-M410 | 14 | 13 | 30 | 24 | 11 | 11 | 13 | 13,14.2 | 9 | 11 | ***1*** |
| | H9 | J2a1*-DYS413<18 | 14 | 13 | 30 | 23 | 10 | 11 | 12 | 15,17.1 | 9 | 11 | ***1*** |
| | H10 | J2a1*-DYS413<18 | 14 | 13 | 31 | 23 | 10 | 11 | 12 | 12,17 | 9 | 11 | ***1*** |
| | H11 | J2a1*-DYS413<18 | 15 | 12 | 28 | 24 | 10 | 11 | 12 | 16,20 | 9 | 11 | ***2*** |
| | H12 | J2a1b-M67 | 14 | 12 | 29 | 24 | 10 | 11 | 12 | 13,18 | 9 | 12 | ***1*** |
| | H13 | J2a1b-M67 | 14 | 13 | 29 | 23 | 10 | 11 | 12 | 13,15 | 9 | 11 | ***1*** |
| | H14 | J2a1b-M67 | 14 | 13 | 29 | 23 | 10 | 11 | 12 | 13,16 | 9 | 10 | ***1*** |
| | H15 | J2a1b-M67 | 15 | 14 | 31 | 23 | 12 | 11 | 12 | 12,13 | 9 | 11 | ***1*** |
| | H16 | J2a1h-DYS445<7 | 15 | 13 | 29 | 23 | 9 | 11 | 12 | 16,16 | 9 | 13 | ***1*** |
| | H17 | J2a1h-DYS445<7 | 15 | 13 | 30 | 24 | 9 | 11 | 12 | 14,19 | 9 | 11 | ***1*** |
| | H18 | J2a1h-DYS445<7 | 16 | 13 | 29 | 23 | 9 | 11 | 12 | 13,16 | 9 | 11 | ***1*** |
| | H19 | J2b*-M12 | 15 | 12 | 28 | 24 | 10 | 11 | 12 | 16,18 | 9 | 12 | ***1*** |
| | H20 | J2b2-M241 | 15 | 12 | 28 | 24 | 10 | 11 | 12 | 15,18 | 9 | 12 | ***1*** |
| *Huelva* | H21 | J1*-M267 | 14 | 13 | 26 | 23 | 10 | 11 | 12 | 14,16 | 10 | 12 | ***1*** |
| | H22 | J1*-M267 | 14 | 13 | 30 | 23 | 10 | 11 | 12 | 13,15 | 10 | 12 | ***1*** |
| | H23 | J1*-M267 | 14 | 13 | 30 | 23 | 10 | 11 | 12 | 13,20 | 10 | 11 | ***1*** |
| | H24 | J1*-M267 | 14 | 13 | 30 | 23 | 10 | 13 | 12 | 13,14 | 10 | 12 | ***1*** |
| | H25 | J2a1*-DYS413<18 | 15 | 13 | 29 | 23 | 10 | 11 | 12 | 12,12 | 9 | 12 | ***1*** |
| | H26 | J2a1b-M67 | 15 | 14 | 30 | 23 | 10 | 11 | 12 | 13,15 | 9 | 11 | ***1*** |
| | H27 | J2a1h-DYS445<7 | 15 | 13 | 29 | 23 | 9 | 11 | 12 | 13,15 | 9 | 11 | ***2*** |
| | H28 | J2a1h-DYS445<7 | 15 | 13 | 29 | 23 | 9 | 11 | 12 | 13,16 | 9 | 11 | ***1*** |
| | H29 | J2b*-M12 | 15 | 13 | 30 | 24 | 10 | 11 | 12 | 14,19 | 9 | 12 | ***1*** |
| | H30 | J2b2-M241 | 15 | 12 | 28 | 24 | 10 | 11 | 12 | 15,18 | 9 | 12 | ***1*** |
| | H31 | J2b2-M241 | 16 | 12 | 28 | 25 | 10 | 11 | 12 | 13,17 | 9 | 12 | ***1*** |


*Fig. 1. Map of Iberia showing the location of Andalusia region. The two studied Andalusian provinces: Granada and Huelva with the municipalities from which blood samples were collected have been highlighted.*

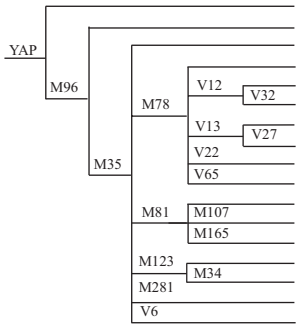
(h=0.8571±0.057). One interesting J2 sub-haplogroup is J2a-M410, with frequencies of 4.80% in Granada compared to 3% in Huelva. High frequencies of the M410 polymorphism have been found in Central and Mediterranean Turkey (19%) as well as in neighbouring Greece (12%)[11,13,34]. Within the J2a-M410 subcluster, we detected the J2a1-DYS413<18, J2a1b-M67 and J2a1h- -DYS445 <7 lineages at frequencies lower than 2%. Comparatively high values of J2a1b have been observed in southern Italy[13] (8.10%), Greece[34] (7.6%) and Turkey[11] (6.31%). In addition, the J2a1h occurs in Mediterranean Anatolia (7%)[44], and is relatively well represented on the island of Crete (4%)[19]. Given the particular phylogeographical pattern of J2a across the Mediterranean as well as in Middle Eastern and western-central Asian populations, some authors have interpreted this scenario as being associated, at least in part, with the spread of agriculture[43]. However, others have proposed that the J2a1b lineage might have emerged in the Aegean area, possibly during the population expansion of the Greek world, including the European coast of the Black Sea[13,45].

Another lineage within the J2 clade is the J2b-M12 (formerly J2e). In Europeans, this has a high prevalence in the Balkans that makes it a specific J subclade of that region[34]. In other surrounding Mediterranean areas, this lineage is present, though it is less common (e.g., Italian Peninsula: 4%), so it has been suggested that it might trace the subsequent diffusion of people of the Balkan region to the west[14]. Interestingly, J2b is virtually absent in other western and central European populations, such as Spanish Basques, Catalans, Sicilians, the Dutch, Hungarians and Slovenians[13,14,34]. The distribution pattern of J2b within Europe is parallel to that observed for E-V13. Both of these account for more than one-fourth of the chromosomes currently found in the southern Balkans, highlighting the strong demographic impact of the expansion in the area, particularly during the Bronze Age[12].

In the Granada sample, a single Y chromosome was also found that exhibited the rarest and basal J* marker, and another chromosome was found corresponding to J2*. Because paragroup J* is very rarely found throughout the Mediterranean area[11,13], this marker should be considered an outstanding feature of the genetic constitution of Granada Andalusians.

**TABLE 3**
FREQUENCY AND THE EXTENT OF DIFFERENTIATION OF Y-CHROMOSOME HAPLOGROUP E OBSERVED BOTH WITH BIALLELIC MARKERS (Y-SNPs) (A) AND MICROSATELLITES (Y-STRs) (B) IN GRANADA ANDALUSIANS

a)




| Lineages[1] | *Andalusian Subpopulations* | |
|---|---|---|
| | *Granada* (n=250) | *Huelva*[2] (n=167) |
| D* | 0 | 0 |
| E* | 0 | 0 |
| E1b1b1* | 0 | 2 (1.20) |
| E1b1b1a* | 0 | 2 (1.20) |
| E1b1b1a1* | 2 ( 0.80)[3] | 1 (0.60) |
| E1b1b1a1b | 0 | 0 |
| E1b1b1a2* | 2 ( 0.80) | 7 (4.19) |
| E1b1b1a2a | 0 | 0 |
| E1b1b1a3 | 0 | 0 |
| E1b1b1a4 | 0 | 1 (0.60) |
| E1b1b1b* | 9 (3.60) | 5 (2.99) |
| E1b1b1b1 | 0 | 0 |
| E1b1b1b2a | 0 | 0 |
| E1b1b1c* | 0 | 0 |
| E1b1b1c1 | 4 (1.60) | 2 (1.20) |
| E1b1b1d | 0 | 0 |
| E1b1b1e | 0 | 0 |
| | 17 (6˙80) | 20 (11.98) |

[1] ISOGG 2010
[2] For comparison purposes are included Hg E data by Ambrosio et al. (2010)
[3] Relative frequencies

b)

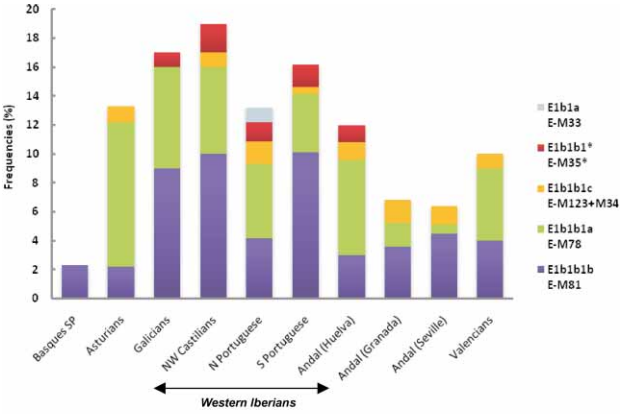
| | Haplotype | Haplogroup | DYS19 | DYS389 I | DYS389 II | DYS390 | DYS391 | DYS392 | DYS393 | DYS385a/b | DYS438 | DYS439 | *N* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Granada* | H1 | E1b1b1a1*- V12 | 13 | 13 | 30 | 23 | 10 | 11 | 13 | 16,16 | 10 | 15 | ***1*** |
| | H2 | E1b1b1a1*- V12 | 13 | 13 | 31 | 24 | 11 | 11 | 13 | 14,16 | 10 | 12 | ***1*** |
| | H3 | E1b1b1a2*- V13 | 13 | 13 | 30 | 24 | 10 | 11 | 13 | 14,18 | 10 | 13 | ***1*** |
| | H4 | E1b1b1a2*- V13 | 13 | 13 | 30 | 24 | 10 | 11 | 13 | 16,18 | 10 | 12 | ***1*** |
| | H5 | E1b1b1b*- M81 | 13 | 14 | 29 | 24 | 9 | 11 | 13 | 13,15 | 10 | 10 | ***1*** |
| | H6 | E1b1b1b*- M81 | 13 | 14 | 30 | 23 | 9 | 11 | 13 | 13,14 | 9 | 10 | ***1*** |
| | H7 | E1b1b1b*- M81 | 13 | 14 | 30 | 24 | 9 | 11 | 13 | 13,13 | 10 | 10 | ***1*** |
| | H8 | E1b1b1b*- M81 | 13 | 14 | 30 | 24 | 9 | 11 | 13 | 13,14 | 10 | 10 | ***2*** |
| | H9 | E1b1b1b*- M81 | 13 | 14 | 30 | 24 | 9 | 11 | 13 | 13,14 | 10 | 11 | ***1*** |
| | H10 | E1b1b1b*- M81 | 13 | 14 | 30 | 24 | 9 | 11 | 13 | 14,14 | 10 | 10 | ***1*** |
| | H11 | E1b1b1b*- M81 | 13 | 14 | 30 | 24 | 9 | 11 | 14 | 13,14 | 10 | 10 | ***1*** |
| | H12 | E1b1b1b*- M81 | 14 | 14 | 30 | 24 | 9 | 11 | 13 | 13,14 | 10 | 11 | ***1*** |
| | H13 | E1b1b1c1- M34 | 13 | 13 | 30 | 23 | 9 | 11 | 14 | 15,15 | 10 | 13 | ***1*** |
| | H14 | E1b1b1c1- M34 | 13 | 13 | 32 | 24 | 10 | 11 | 13 | 16,20 | 10 | 12 | ***1*** |
| | H15 | E1b1b1c1- M34 | 14 | 14 | 32 | 23 | 10 | 11 | 13 | 16,18 | 10 | 13 | ***1*** |
| | H16 | E1b1b1c1- M34 | 15 | 14 | 31 | 24 | 11 | 11 | 13 | 16,17 | 10 | 11 | ***1*** |

Table 3 shows the frequency and composition of Hg E-M35 (formerly E3b) among Andalusians from Granada; for comparison, the Hg E results we observed in the population of Huelva[15] are also presented. The frequency of Hg E and its internal diversity are lower in Granada than in Huelva. In the Granada sample, Hg E is significantly represented by the E-M81 subclade; that is, more than half (9 occurrences) of the Y-chromosomes containing the YAP element (YAP+) carried the mutation M81 (3.60% of the total; 53% of the Hg E) being identified all of them as E-M81* (E1b1b1b*) because these chromosomes did not carry derivate alleles at M107 and M165. Other authors[31,32,46] found frequencies of M81 in Iberia ranging from 0 to 9%. Basques from Spain register the lowest levels (1–3.6%) of Hg E[14,32,47,48].

The other Hg E subclades observed in the Granada sample correspond to E-M78 and E-M123, the latter occurring at high frequencies (10–12%) among Jews[49]. The frequency of M78 is rather low (1.60%), and only two major internal branches of this haplogroup, E-V12 and E-V13, have been found. Thus, we can conclude that Granada is less diverse than Huelva for Hg E; both Andalusian subpopulations contain an almost-equal frequency of E-M81 (≈3%) and E-M123 (M34) (≈1.40%), whereas the V13 lineage is observed to be four times more common in Huelva than in Granada. Consequently, the estimated Hg E diversity (h=0.676±0.09) was low in the Granada sample, in contrast to what was found (0.821±0.06) at the other extreme side of the region. In support of this, the value $\chi^2$=3.32, p=0.069 was highly correspondent to the $F_{ST}$=0.083, p=0.055, showing an almost significant genetic differentiation in the geographic distribution pattern for Hg E in the region.

For the sake of comparisons, the composition of E clade in some selected Iberian populations from different regions (most sample sizes ≥50) is shown in Figure 2. This simple plot interestingly highlights the remarkable heterogeneous occurrence of the E-M81 and E-M78 subclades among Spanish and Portuguese populations, as well as the widespread distribution of E-M123 (M34) within the Peninsula. Likewise, it is interesting to note that the M35* paragroup is concentrated in western Iberian coastal populations, though it has not been found in other southern and eastern Spanish samples (e.g., Granada, Valencia). In most Mediterranean populations, the M33 and M2 mutations show frequencies lower than 1%. Analysis of the distribution patterns of some lineages within the Peninsula (e.g., E-M81 and E-V13) in comparison to historical records suggests that further confirmation of these results is required, giving particular attention to adequate sample sizes, sampling processes within local areas and autochthony of the people sampled.

A Hierarchical Cluster Analysis was performed (Figure 3) to examine the population structure based on five J haplogroup/paragroup frequencies for a set of worldwide populations (Table 1). The HCA was constructed on the three first principal components (PCs) which account for a high percentage of the variance (98%), with PC1 capturing most of variation (78%). The multivariate




(a) Underhill et al. 2000; (b) Cruciani et al. 2004; (c) Semino et al. 2004; (d) Alonso et al. 2005; (e) Adams et al. 2008; (f) Flores et al. 2004; (g) Ambrosio et al. 2010; (h) Present study.

*Fig. 2. Phylogenetic diversity of Y-chromosome haplogroup E among some Iberian populations from Spain and Portugal. Occurrences for haplogroup E/total sample sizes and sources for each population are as follow: Basques SP 5/222 (a, b, c, d); Asturians 12/90 (b); Galicians 15/88 (e); NW Castilians 19/100 (e); N Portuguese 32/219 (b, e, f); S Portuguese 22/127 (b, e); Andal (Huelva) 20/167 (g); Andal (Granada) 17/250 (h); Andal (Seville) 11/155 (f); Valencians 7/73 (e).*

analysis provided five clusters, with the J1-M267, J2* and J2a-M410 lineages significantly defining the genetic map. When the inertia decomposition on the first three axes is computed, the quotient (inertia interclusters/inertia total) equals 0.88. This value is highly coherent with the number of major ramifications shown by the tree, and it demonstrates that a high percentage of the data variation is explained by these five clusters. Cluster 1 (C1) is positioned on the positive sides of axes 1 and 2, and it is shaped by 24 populations, 87% of which are of European origin (including the two Andalusia samples studied), plus two other with African origins: Sudanese and Ethiopians (Oromo)[14]. C1 is defined by low and relatively homogeneous frequencies of the active variables J2*, J1-M267 and J2a-M410, each of which show significant differences with respect to the overall mean (p< 0.004). PC3 is determined by the J2* variable which, in turn, genetically characterises C2 with high mean frequencies of this J paragroup (cluster mean=13.35; p= 0.000). Consequently, C2 is better defined by the plane PC1-PC3, and includes Italians from Sicily and southern Italy, Turks and Ashkenazi and Sephardic Jews. Likewise, C3 includes samples from mainland Greece and Crete, Albania, Caucasus, Syria, Iran, and Egypt. C3 is defined by highly significant frequencies of J2a-M410 (cluster mean=17.15; p=0.000).

Clusters C4 and C5, which are positioned across the negative side of axis 1, are both genetically characterised by high levels of the J1-M267 haplogroup. C5, which contains the highest mean frequency of this lineage (cluster mean=65.46; p=0.000), is only influenced by Yemen and Qatar in the Arabian Peninsula; both populations occupy extreme outlier positions. Interestingly, C4, which has an intermediate location between C5 and the centroid, in-

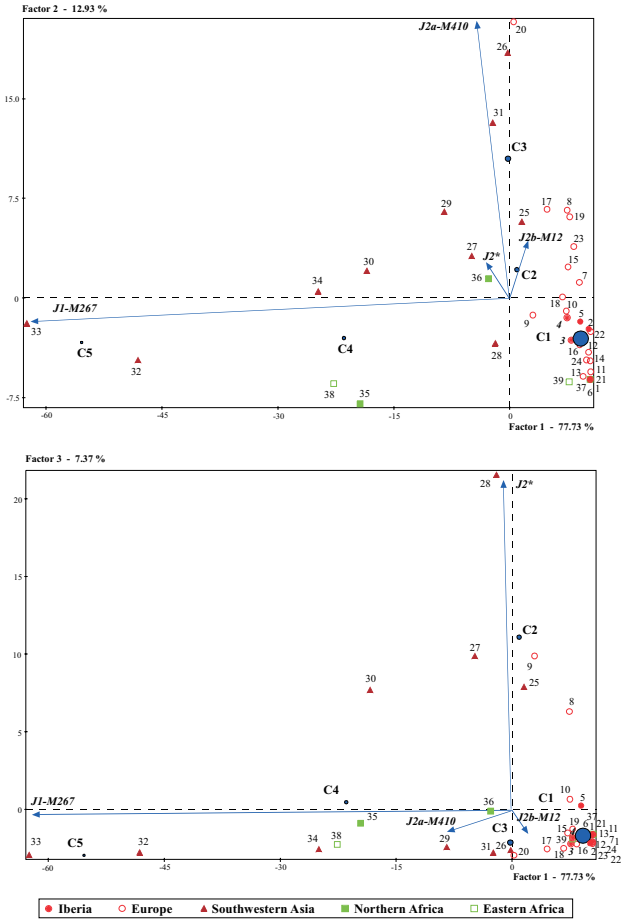



*Fig. 3. Hierarchical Cluster Analysis based upon five haplogroup J lineages: J*, J1-M267, J2*, J2a-M410 and J2b-M12 in a set of Mediterranean and other surrounding populations. Two plots are displayed: a) defined by PC1-PC2 and b) defined by PC1-PC3. Arrows (vectors) show the positions of the normed eigenvectors of the subhaplogroup variables. Dots show the positions of cluster's centers after consolidation, with circle sizes proportional to the number of populations that falls in each cluster.*

cludes two Near East populations, Iraq and the United Arab Emirates, and another two African samples: one from Tunisia and another from Ethiopia (Amhara). It appears that haplogroup J1-M267 could have originated in the south of the Arabian Peninsula (>70% in Yemen) and then spread to other geographically contiguous countries, making it representative of the southern Arab tribes. The presence of J1 chromosomes in Tunisians and Ethiopians could be a result of the Islamic expansion. Tunisia was an important region for the expansion of Islam in North Africa and South Europe.

The Granada sample has been added to the HCA of 74 worldwide populations of Hg E and its lineages that is shown in[15]. Granada is grouped in cluster C1, where most of the European populations are included, while the Huelva population is positioned in cluster C2. The main reason for this is the difference between the numbers of cases of the E-V13 lineage observed between the two Andalusian samples. All the other features of the new HCA (not shown here) did not change with respect to those that had been presented in our early paper, and we refer to the results and discussion therein.

### *Internal haplotype diversities*

Based on 15 Y-STRs, we found 160/179 and 153/167 different »unique« Y-chromosome STR haplotypes among Andalusian people of Granada and Huelva, respectively. Table 2b presents the haplotypes assigned to Hg J in the studied Andalusian samples. The intra-haplogroup haplotype diversity of Hg J, based on extended haplotypes of 9 loci was rather high: H=0.972±0.020 (Granada) and H=0.954±0.057 (Huelva) with mean variances of 0.443 and 0.459, respectively. When considering the minimum haplotype of 7 loci, the corresponding *H* values were, respectively, 0.936±0.033 and 0.945±0.059. The modal haplotype structure in the Andalusian J pool across 9 loci was DYS19(14)/DYS389I(13)/DYS389II(16)/DYS390(23)/DYS391(10)/DYS392(11)/DYS393(12)/DYS438(9)/DYS439(12), which is present on both sides of the region. Both the observed modal and its surrounded haplotypes contain the so-called Cohen modal haplotype [DYS19(14)/DYS390(23)/DYS392(11)/DYS393(12)] without considering DYS388 microsatellite[50]. This haplotype is highly represented in the Middle-Eastern J pool[11,51]. In our Andalusian samples, we found that 11 out of 35 males carry these four allele motifs and that more than half of those Y-chromosomes were, in turn, associated with the J1 haplogroup; the other 4 chromosomes belonged to the J2a lineages. Interestingly, the 4 Andalusians from Huelva bearing a M267 mutation carried the Cohen haplotype, and three of them were also associated with the haplotype 14/13/17/23/10/11/12 (the J modal haplotype in Huelva). This allelic association has not been detected among the J1 of the 7 Y chromosomes from Granada. Nevertheless, the J1-14/13/13 (26)/23/10/11/12 lineage is present in both sampled Andalusian subpopulations; this haplotype is considered »rare« because of the presence of allele 26 at the DYS389II microsatellite. Rare haplotypes are generally young in an evolutionary sense, so that they represent powerful markers for recent movements through space of individuals or populations[52].

Notably, three out of five Y-chromosomes (from the total studied Andalusia sample) carrying the M12 (or its derived allele M241) are associated with the haplotype 15/12/16/24/10/11/12. The ancestral paragroup J* that was detected in one individual from Granada was also associated to the same haplotype.

The estimated haplotype diversities (extended haplotype) for the J1 and J2 subclades were, respectively: 0.927±0.066 (J1) and 0.972±0.022 (J2); when Granada and Huelva data sets were analysed separately, the H values differed only slightly (data not shown).

A total of 16 different Hg E STR haplotypes have been observed among 17 YAP (+) Y-chromosomes (H=0.992±0.023) (see Table 3b). This contrasts with the rather moderate haplogroup diversity we observed in our Granada sample (h=0.676±0.094). The modal haplotype ob-

served was 13/14/17/24/9/11/13/13.14/10/10. It is worth noting the frequent presence of DYS19*13 (88%) (h= 0.404), which seems to define a particular affinity of the specific alleles associated with YAP (+) males. In contrast, the fairly dominant E-M81 haplogroup among Andalusians from Granada showed a rather low haplotype diversity of 0.833±0.126 (0.722±0.160 based on minimum haplotype) and a low mean Y-STR variance (0.0833) in allele size. Within the E-M81 cluster, the occurrence of the so-called Maghrebin haplotype 13/14/16/24/9/11/13 is predominant (5/9 Y-chromosomes M81). The frequency of this haplotype in North Africa is appreciable[10]. Furthermore, the two eastern Andalusian E-V13 chromosomes detected are associated with the same seven loci, with haplotype 13/13/17/24/10/11/13, that is especially represented in the Balkan region, and in Greece and Crete, in particular. These genetic particularities are also present in the western side of the region (Huelva province) though with heterogeneous E-V13 and E-M81 frequency patterns (see for details[15,40]).

The phylogenetic network of J2a-M410 and J1-M267 STR haplotypes is shown in Figure 4, and the frequency of surface distributions (surface maps) of these two lineages across the Mediterranean area is shown in Figure 5. The J2a-M410 network was constructed, for a total of 128 males, from 11 Mediterranean population samples yielding a total of 92 different haplotypes (71% could be placed on the network). Twenty-two haplotypes were detected more than once, and they correspond to 79 individuals. No haplotype was common to all eleven populations.

Haplotype age is expected to be proportional to the number of connections with other haplotypes. From the node associated with modal haplotype 14/13/16/23/10/11/12, which is located in the core of the network, there emerge six adjacent haplotypes, most of which are shared by two or more populations. The putative »ancestral« node is represented by Turks, Sicilians and Andalusians from Granada.

The J2a network, within which most of the links are the expression of only one single-step mutation, could be structured into three informative subclusters. The right-hand subcluster mostly concentrates Sicilian microsatellite haplotypes (47%). Its largest node is located on the main branch. It is distant from the modal haplotype and comprises 6 occurrences of the most common J2a Sicilian haplotype 15/13/16/23/9/11/12, and it is shared, to a lesser extent, with Andalusian and Balkan chromosomes. All the Huelva and most of the Granada haplotypes appear to be mainly related with Balkan, Turk and Caucasian single haplotypes. On the left side of the network, we found the less diversified subcluster, with the main branch (lineage) being occupied by a high proportion of the Turk -and Caucasian- related haplotypes; the punctuated presence of Sicilian, Greek and Granada single haplotypes, which are depicted as peripheral branches, constitute another interesting feature of this subcluster. The third clustering of haplotypes includes the largest node found in the network (n=11), with 4 different populations from Turkey, the Caucasus and a lower number from Greece and Granada, where the shared haplotype is 14/13/17/23/10/11/12.

The network of J1 haplotypes (in Figure 4) was based on 294 individuals from 6 population data sets. The network is less complex than that of the J2a-M410 subclade, possibly due to the more restricted geographical pattern of the J1-lineage, with peak frequencies in the eastern

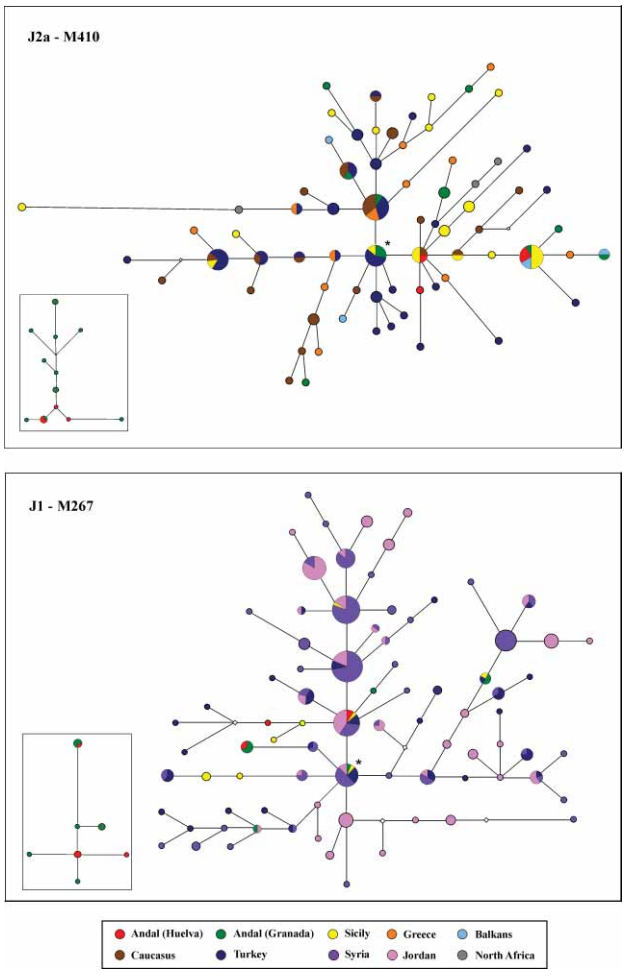



*Fig. 4. Median-joining microsatellite haplotype networks (shortest tree) for J2a-M410 and J1-M267 haplogroups based on 7 Y-STRs loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393). Circles are proportional to the number of individuals sharing that haplotype. Populations used to construct the networks have been taken from the literature and the present study: Tunisians, Egyptians from Northern and Southern, Arabs and Berbers from Algeria (Arredi et al 2004); Turkish (Cinnogliu et al 2004); Sicily (Di Gaetano et al 2009); Balkan region and Greece and Caucasus (Battaglia et al 2008); Syria and Jordan (El-Sibai et al 2009); and Andalusians (present study). Occurrences for each lineage and populations have been as follow: J2a-M410 (Andalusians from Granada, 12/250 and Huelva, 5/167; Sicily, 23/236; Balkans, 5/255; Greece, 13/92; Turkey, 42/523; Caucasus region, 25/104; Egypt, 1/73; Tunisia, 1/148, Arabs (Algeria), 1/35), J1-M267 (Andalusians from Granada, 7/250 and Huelva, 4/167; Sicily, 9/236; Turkey, 44/523; Syria, 138/356; Jordan, 92/273).*

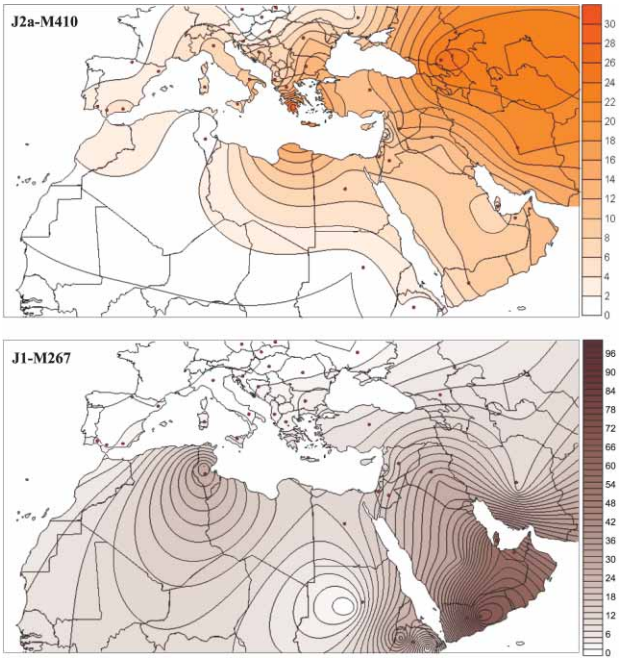



*Fig. 5. Surface maps showing the genetic patterns of J1-M267 and J2a-M410 haplogroups across Mediterranean and other neighbour geographical areas. The maps were constructed from the data of the 39 populations in Table 1.*

end of the Mediterranean region. The ancestral node is defined by the 14/13/17/23/10/11/12 modal haplotype, which is significantly represented by Syrian, Jordanian and Turkish samples. The network displays a dominant main branch comprising three single-step linked large nodes, which contains most Syrian (56/138) and neighbouring Jordanian samples[41] and some Andalusian samples from Huelva. Most of the Andalusian haplotypes are positioned close to the ancestral node. Only one haplotype was present in both eastern and western Andalusia.

**TABLE 4**
THE AGE ESTIMATES (IN YEARS) OF THE MOST RECENT COMMON ANCESTORS (TMRCA) FOR MUTATIONS DEFINING HAPLOGROUPS/SUBHAPLOGROUPS J AND E IN SOUTHERN SPAIN

| | | TMRCA (CI)[2] |
|---|---|---|
| Haplogroup/ Sub-haplogroups age[1] | N | Andalusians |
| J-12f2 | 35 | 9193 (8663–9723) |
| J1-M267 | 11 | 5184 (4852–5515) |
| J2-M172 | 23 | 9105 (8477–9553) |
| J2a-M410 | 17 | 7019 (6603–7435) |
| J2b-M12 | 5 | 4126 (3842–4409) |
| E-M35 | 37 | 8758 (8220–9297) |
| E1b1b1a-M78 | 14 | 7165 (6723–7607) |
| E1b1b1a2-V13 | 9 | 3493 (3265–3721) |
| E1b1b1b-M81 | 13 | 2699 (2512–2887) |
| E1b1b1c1-M123 (34) | 6 | 6215 (5833–6597) |

[1] Values were computed over 15 loci microsatellites
[2] CI, is the 95% Confident Interval

The time to most recent common ancestor (TMRCA) of the Andalusian haplogroups was calculated using BATWING, and it is depicted in Table 4. The range varies between 2.1 and 9.7 ky, and hence, they are posterior to the LGM and, in general, are included in the Neolithic or Bronze Age. The TMRCAs estimated by[7] are 52,500 years ago for the E haplogroup, 47,500 years ago for the E1b1 subhaplogroup and 38,500 years ago for the IJ major clade. Considering these values, it is most likely that the J and E lineages were introduced into the Iberian Peninsula during or after protohistory, with low diversification there.

## Discussion

Determining the effects of population mixing requires knowledge of the genetic characteristics of the migrants and their relationship to the host population, the regions of origin of those migrants, the time of the migrations, the proportion of immigrants in the host population, the areas in the host region where they settled and the subsequent internal movements of the mixed populations. The historic sources that describe the magnitude and composition of the migrations towards the Iberian Peninsula, despite being highly numerous, are very far from providing enough data to evaluate all of these variables. However, the effort required for this evaluation is necessary to provide a coherent explanation of the obtained genetic results, in which the documented migrations should not be ignored because of uncertainty about their characteristics and the complexity of the analysis.

The sub-haplogroups of the J and E clades of the Y chromosome do not allow for clear distinction among the contributions of the paternal lineages that originated in nearby populations, although there are important differences among their frequencies. In this context, the territories of origin of the emigrants that left their signature in the populations of the Peninsula can only be defined in broad terms. For example, we cannot distinguish, with the current data, among the different contributions of the northern and southern Arabs for J1-M267 lineage.

The presence of the J and E haplogroups can be mainly explained as a result of the migrations that have taken place since the Neolithic and, in particular, the ones that occurred during protohistory and history up to the modern age. The observed J1 sub-haplogroup in the Iberian Peninsula is most likely a consequence of the migration of Arabic populations during the Islamic expansion, while the J2 sub-haplogroup seems to be related to the Greek and Phoenician colonies that were established in the Peninsula due to commerce that was mainly linked to the mining-related wealth on that territory and the accompanying maritime trades. In both cases, these movements were likely highly directional, and new immigrants settled only in some regions of the Peninsula territory.

The Phoenicians settled in colonies on the coast of Granada. Nevertheless, commerce was more intense with Huelva, where the Tartessos kingdom was established, and which maintained the strongest bonds with the Greeks[53]. The ancient alphabets used in Huelva and Granada were different, which indicates distinct cultural, and possibly genetic influences among the migrations that introduced them. The Carthaginians were made up, in part, of Phoenicians who aggregated with other groups that would have contained Berbers from the area surrounding Carthage. Their armies were formed by mercenaries from the Mediterranean region, including Iberia. The efforts to quantify the contributions of these migrations are currently idle.

Different processes could explain the presence of the sub-haplogroups E-M81 and E-M78; the first of which is associated with the Berber component of the Islamic expansion, and to a lesser extent, with the older expansion, whereas the second would be associated with the previously mentioned increases in trade, particularly the migrations directed to the Tartessos kingdom.

With respect to the migrations associated with the Muslim conquest and the Christian reconquest of the Iberian Peninsula, it is first important to point out the fundamental military character of the Muslim expansion, which involved an army organised around a tribal base. This army entered the Iberian Peninsula in the year 711, and it was composed of northern and southern Arabs and Berbers. The subsequent fights among the diverse invading groups were numerous, which necessitated the entry of a new army in the year 741, made up of contingents from the diverse Islamised regions in the near East and Egypt. These contingents were separately settled in the Peninsula. The contingent of Damascus settled in Granada, that from Jordan in the province of Malaga, and the one from Egypt close to Huelva. The Arabs settled preferably in the cities, while the Berbers mostly inhabited the mountain regions, reaching the centre and northwest of the Peninsula, including Galicia, the north of Portugal and León[54]. For several centuries, the northeast was also occupied by the Muslims, extending to Barcelona and Pamplona. Nevertheless, the main settlements were in Andalusia, where the capital of the caliphate, Córdoba, was located. Córdoba was a great city at that time.

Al-Andalus, the name given by the Muslim conquerors to the Iberian Peninsula, was soon isolated from the rest of the Islamic Empire; however, two other Berber armies later entered Iberia: the almorávides (1090–1147) and the almohades (1146–1229). The powerful Cordoban State of the Omeyas disappeared in 1031, leaving the former kingdom a divided territory in several independent kingdoms or Taifas. This favoured the progressive Christian reconquest, which was completed in 1264, with the exception of Granada, which was reconquered in 1492.

In the VIII century, the population of Spain was estimated to be between 7 and 8 million inhabitants and, during that time, approximately 150,000 to 200,000 Muslim warriors entered the territory[55]. The largest part of that immigrant population was north African (120,000 to 160,000) of Berber origin, which represented between the 1.5 and 2% of the Peninsula population. The corresponding population of Arabs is estimated to have been between 30,000 and 40,000 (0.4–0.5%), and approximately half of them were southern Arabs or Yemenis. It is unknown how many Muslims of those who entered the Iberian Peninsula in the VIII century ended up settling there, how many died because of the numerous conflicts, or how many brought their families to the Peninsula following the conflicts. The previously determined relative values for genetic contribution based on data from the Y-chromosome should be multiplied by at least 2 because all these of immigrants were males, and they should probably be multiplied by 2 or 3 because most of them were of reproductive age. Hence, in the case of the Berbers, the percentage of genes of the Y chromosome introduced would be approximately 10%, and it would be approximately 2% for Arabs, if all of these individuals replaced resident males in the transmission of their Y chromosomes to the next generation. In the less-populated regions of the Iberian Peninsula, those percentages could be even higher. Nevertheless, that scenario seems less feasible and too extreme, and that is why we propose relative intermediate values, with a percentage of 5% for the Berbers and 1% for the Arabs for the total in the Peninsula.

The genetic effect of the immigrants in the host population would have also been enhanced by polygamy, a marital behaviour that was practised preferably by the wealthiest individuals, who, to a great extent, were Arabs. Nevertheless, this effect would have been very different in the context of the whole genome, and in the case of the Y chromosome in particular. The 10 Emirs and Caliphes who made up the Umayyad dynasty in Al-Andalus were all sons from concubine slaves, almost all of whom were Spanish and from the North of the Peninsula. The founder of this dynasty, Abd al-Rahmân I, was the son of a Berber woman, and his son and Emir successor had a Spanish mother. Therefore, the genome of Hišâm II, the tenth and last Caliphe of the Umayyad dynasty, would have mostly originated from the Iberian Peninsula and would not be more than $0.5^{10}=0.001$ of Arab descent, although the Y chromosome would still be of fully Arab origin[55].

The development of the Christian reconquest caused complex episodes of repopulation, colonisation and population assimilation; in other words, it was a series of historical events that could have strongly influenced the spatial genetic variation patterns inside the Peninsula. In this period, Andalusia was repopulated with »old Christian« peasants from Castile and Aragón, as well as those from more distant areas such as Galicia and even Portugal.

After the conquest of the Granada kingdom in 1492, numerous Moriscos (descendants of the Muslim population) remained in the region. Finally, in 1609, the expulsion of the Moriscos from Spain was decreed. Their numbers have been estimated at between 300,000 and 500,000,

and many of them were sent to the Maghreb[56]. Nevertheless, a considerable percentage of the Moors remained in Spain because their professions were very hard to replace. Subsequently, some of the expelled Moriscos, or their descendants, were able to return to Spain. Whether the genetic characteristics of the expelled Moriscos who returned notably differed from those subpopulations belonging to the regions where they lived is an interesting question because, after the 900 years that passed from the Muslim invasion to their expulsion, it is highly improbable that there remained a significant correlation between religion and genetics.

The number of Jews that entered Spain is unknown, and although some sources have suggested their presence extends back long before the Roman Empire, the most credible hypothesis is that they settled there during the time of the Roman Empire and that their numbers increased later with the Muslim invasion. In the city of Elvira (close to present-day Granada city), they were numerous, and their presence there is known since before the fourth century due to writings from bishops from that city who prohibited the Christians from allowing their daughters to marry Jewish people[57]. In 1492, the expulsion from Spain of those Jews who did not convert to Christianity was decreed. The number of expelled Jews is unknown. Nevertheless it has been proposed[58] the following distribution: during the end of the 15th century, the total number of Jews in Spain was around 400,000 for a total population of 7–8 million, from which 240,000 were likely converted, and the remaining 160,000 would have accepted exile. Hence, the Jewish population was estimated to constitute 3.5% of the total demography of the country after their expulsion, half of which would have been males, and a portion of those, which could have been considerable, were not ethnic Jews. The largest portion of the expelled Jews ended up in Portugal, which had a population of 1–1.5 million. The size of the Jewish population in Portugal in 1492 has been estimated to be 190,000 which represents the sum of all of the individuals who lived there before and those who were expelled from Spain. During the period of the political union of Spain and Portugal (1581–1642), some of those Jews returned to Spain. All of these numbers contain a large degree of uncertainty, but they do demonstrate the enormous disparity between the frequency of Jews in Spain and Portugal and the high dependence of the demography of the host population on the impact of this kind of migration in the correspondent gene frequencies. The Jewish people who are relevant to the study of migrations in the Iberian Peninsula are those who remained in Spain and whose descendants still live there, independent of their religion and considering only the origin of their ancestors. The current Sephardic Jews, who are only partially descendants of the Jews who lived on the Peninsula[59] and who are mixed with individuals from countries where they lived after their expulsion from Spain are of secondary interest for this study, except as comparative populations. A difficulty often found in inferring Jewish migrations from genetic information is that they do not have clearly dominant haplogroups or sub-haplogroups in the Y chromosome, which makes the effect of admixture with other populations weakly reflected in their genomes.

With respect to the Romans and the Visigoths, Romans appear not differ much genetically from the Iberians, particularly in relation to the Y chromosome, such that the gene frequencies of the population would not have changed much due to these invaders. The Visigoths, who invaded in 409, were between 80,000 and 100,000 individuals, of which 20,000 were warriors. They settled in the middle and north of the Peninsula, and for several generations, they married among each other. The size of their masculine population is hence modest, and marriage inside the group would have reduced their mixing with the general population. Nevertheless, their concentration in a reduced and well-defined area of the Peninsula could have left some traces that persisted until the present. The result of the concentration of immigrants in a part of the territory of a country and the subsequent migration of a portion of that population to other regions of the country after several generations, leads to relatively lower frequencies of their genetic markers in the part of the territory that received the emigrant genes in a second phase, compared to an initial homogeneous distribution of emigrants around the whole territory.

The results that have been obtained using diverse methods coincide in several aspects, and the frequencies observed in Andalusia can be adequately explained by taking into consideration the frequencies of those haplogroups in the population, the origin of the migrations and their estimated sizes. The HCA of the J haplogroup has established groupings of populations that locate the two Andalusian samples in the group constituted by the majority of the European populations, and it establishes significant differences between those and the Arabic populations for the J1-M267, and the populations of Greece, Iran and the Caucasus for the J2a-M410. The HCA for the E haplogroups is placed in two different clusters in Huelva and Granada, due mainly to the highest frequency of the E-V13 lineage in Huelva, which is most likely associated with the intensity of exchanges with the Tartessos kingdom. The contour maps also suggest the origin in the south of Arabia for J1-M267 lineage. These maps indicate movement across the Mediterranean and across the extensive and less populated North African region. A demic diffusion model for these movements seems unlikely. A population movement passing central Europe and reaching the Iberian Peninsula is not supported by these maps. The network, which could only be constructed from a lower number of populations compared with the other methods, because of the lack of appropriate data in the literature, locate several of the Andalusian haplotypes in the most central regions of these networks. This seems to indicate a movement without intermediate phases, in which the immigrating populations were sources of those sub-haplogroups until they reached the Peninsula. The estimated ages of microsatellite variation associated with haplogroups/subhaplo-

groups J and E in Andalusia do not support an introduction of their different lineages prior to the Neolithic.

In conclusion, the analysis of the Y chromosome was conducted through 2 of its haplogroups (J and E), which are some of the most mutationally diverse and most interesting haplogroups for the study of the peopling of the Mediterranean area. The selection of a geographic region (Iberia) situated in one of the extremes of this geographical space, through which many highly diverse migratory populations entered, and which we carried out using appropriately selected samples, allows a precise understanding of the global result of these migrations until the present time. The size, origin, timing and mode of introduction of migrants and the genetic characteristics of the studied population samples, which are critical variables for these studies, have been discussed here in the context of the Andalusia region. Our findings from this survey highlight the complexity of the migration processes that have taken place throughout the Mediterranean, which were highly relevant from protohistorical times, and demonstrate the impact of the different population sources on the current genetic composition of the population of Spain.

## Acknowledgements

We gratefully acknowledge all doctors, nurses and Andalusian donors who generously participated in this research. In addition, we are indebted to Dr. A. Fernández-Montoya and Dr. D. Prados, Directors of the Blood Transfusion Centres of Granada and Huelva, respectively, as well as to Dr. Antonio Fernández-Jurado, Head of the Haematology Service of the Juan Ramón Jiménez Hospital in Huelva. Financial support has been provided by the Spanish Ministry of Science and Innovation: Research Project CGL2006-04749/BOS (to RC) and by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca: MIUR-PRIN 2007 (to AN).

## REFERENCES

1. BIRABEN J, MASSET C, THILLAUD P, Le peuplement préhistorique de l'Europe. In: Histoire des Populations de l'Europe (Libraire Arthème Fayard, Paris, 1997). — 2. ALMAGRO-BASCH M, DEL AMO M, BELTRÁN A, Huelva: Prehistoria y Antigüedad (Editorial Nacional, Madrid, 1974). — 3. SIMONET F, Descripción del reino de Granada bajo la dominación de los naseritas (Ediciones Extramuros, Sevilla, 1860). — 4. CALDERON R, AMBROSIO B, GUITARD E, GONZALEZ-MARTIN A, ARESTI U, DUGOUJON JM, Hum Biol, 78 (2006) 663. — 5. CHIARONI J, UNDERHILL PA, CAVALLI-SFORZA LL, Proc Natl Acad Sci USA, 106 (2009) 20174. — 6. HAMMER MF, ZEGURA SL, Annu Rev Anthropol, 31 (2003) 303. — 7. KARAFET T, MENDEZ FL, MEILERMAN MB, UNDERHILL PA, ZEGURA SL, HAMMER MF, Genome Res, 18 (2008) 830. — 8. UNDERHILL PA, KIVISILD T, Annu Rev Genet, 41 (2007) 539. — 9. UNDERHILL PA, MYRES NM, ROOTSI S, METSPALU M, ZHIVOTOVSKY LA, KING RJ, LIN AA, CHOW CE, SEMINO O, BATTAGLIA V, KUTUEV I, JARVE M, CHAUBEY G, AYUB Q, MOHYUDDIN A, MEHDI SQ, SENGUPTA S, ROGAEV EI, KHUSNUTDINOVA EK, PSHENICHNOV A, BALANOVSKY O, BALANOVSKA E, JERAN N, AUGUSTIN DH, BALDOVIC M, HERRERA RJ, THANGARAJ K, SINGH V, SINGH L, MAJUMDER P, RUDAN P, PRIMORAC D, VILLEMS R, KIVISILD T, Eur J Hum Genet, 18 (2010) 479. — 10. ARREDI B, POLONI ES, PARACCHINI S, ZERJAL T, FATHALLAH DM, MAKRELOUF M, PASCALI VL, NOVELLETTO A, TYLER-SMITH C, Am J Hum Genet, 75 (2004) 338. — 11. CINNIOGLU C, KING R, KIVISILD T, KALFOGLU E, ATASOY S, CAVALLERI GL, LILLIE AS, ROSEMAN CC, LIN AA, PRINCE K, OEFNER PJ, SHEN P, SEMINO O, CAVALLI-SFORZA LL, UNDERHILL PA, Hum Genet, 114 (2004) 127. — 12. CRUCIANI F, LA FRATTA R, TROMBETTA B, SANTOLAMAZZA P, SELLITTO D, COLOMB EB, DUGOUJON JM, CRIVELLARO F, BENINCASA T, PASCONE R, MORAL P, WATSON E, MELEGH B, BARBUJANI G, FUSELLI S, VONA G, ZAGRADISNIK B, ASSUM G, BRDICKA R, KOZLOV AI, EFREMOV GD, COPPA A, NOVELLETTO A, SCOZZARI R, Mol Biol Evol, 24 (2007) 1300. — 13. DI GIACOMO F, LUCA F, POPA LO, AKAR N, ANAGNOU N, BANYKO J, BRDICKA R, BARBUJANI G, PAPOLA F, CIAVARELLA G, CUCCI F, DI STASI L, GAVRILA L, KERIMOVA MG, KOVATCHEV D, KOZLOV AI, LOUTRADIS A, MANDARINO V, MAMMI C, MICHALODIMITRAKIS EN, PAOLI G, PAPPA KI, PEDICINI G, TERRENATO L, TOFANELLI S, MALASPINA P, NOVELLETTO A, Hum Genet, 115 (2004) 357. — 14. SEMINO O, MAGRI C, BENUZZI G, LIN AA, AL-ZAHERY N, BATTAGLIA V, MACCIONI L, TRIANTAPHYLLIDIS C, SHEN P, OEFNER PJ, ZHIVOTOVSKY LA, KING R, TORRONI A, CAVALLI-SFORZA LL, UNDERHILL PA, SANTACHIARA-BENERECETTI AS, Am J Hum Genet, 74 (2004) 1023. — 15. AMBROSIO B, DUGOUJON JM, HERNANDEZ C, DE LA FUENTE D, GONZALEZ-MARTIN A, FORTES-LIMA CA, NOVELLETTO A, RODRIGUEZ JN, CALDERON R, Ann Hum Biol, 37 (2010) 86. — 16. HAMMER MF, SPURDLE AB, KARAFET T, BONNER MR, WOOD ET, NOVELLETTO A, MALASPINA P, MITCHELL RJ, HORAI S, JENKINS T, ZEGURA SL, Genetics, 145 (1997) 787. — 17. ROSSER ZH, ZERJAL T, HURLES ME, ADOJAAN M, ALAVANTIC D, AMORIM A, AMOS W, ARMENTEROS M, ARROYO E, BARBUJANI G, BECKMAN G, BECKMAN L, BERTRANPETIT J, BOSCH E, BRADLEY DG, BREDE G, COOPER G, CORTE-REAL HB, DE KNIJFF P, DECORTE R, DUBROVA YE, EVGRAFOV O, GILISSEN A, GLISIC S, GOLGE M, HILL EW, JEZIOROWSKA A, KALAYDJIEVA L, KAYSER M, KIVISILD T, KRAVCHENKO SA, KRUMINA A, KUCINSKAS V, LAVINHA J, LIVSHITS LA, MALASPINA P, MARIA S, MCELREAVEY K, MEITINGER TA, MIKELSAAR AV, MITCHELL RJ, NAFA K, NICHOLSON J, NORBY S, PANDYA A, PARIK J, PATSALIS PC, PEREIRA L, PETERLIN B, PIELBERG G, PRATA MJ, PREVIDERE C, ROEWER L, ROOTSI S, RUBINSZTEIN DC, SAILLARD J, SANTOS FR, STEFANESCU G, SYKES BC, TOLUN A, VILLEMS R, TYLER-SMITH C, JOBLING MA, Am J Hum Genet, 67 (2000) 1526. — 18. UNDERHILL PA, PASSARINO G, LIN AA, SHEN P, MIRAZON LAHR M, FOLEY RA, OEFNER PJ, CAVALLI-SFORZA LL, Ann Hum Genet, 65 (2001) 43. — 19. KING R, OZCAN SS, CARTER T, KALFOGLU E, ATASOY S, TRIANTAPHYLLIDIS C, KOUVATSI A, LIN AA, CHOW CET, ZHIVOTOVSKY LA, MICHALODIMITRAKIS M, UNDERHILL PA, Ann Hum Genet, 72 (2008) 205. — 20. MALASPINA P, CIMINELLI BM, VIGGIANO L, JODICE C, CRUCIANI F, SANTOLAMAZZA P, SELLITTO D, SCOZZARI R, TERRENATO L, ROCCHI M, NOVELLETTO A, J Mol Evol, 44 (1997) 652. — 21. HANSON EK, LUBENOW H, BALLANTYNE J, Anal Biochem, 387 (2009) 303. — 22. GILL P, BRENNER C, BRINKMANN B, BUDOWLE B, CARRACEDO A, JOBLING MA, DE KNIJFF P, KAYSER M, KRAWCZAK M, MAYR WR, MORLING N, OLAISEN B, PASCALI V, PRINZ M, ROEWER L, SCHNEIDER PM, SAJANTILA A, TYLER-SMITH C, Forensic Sci Int, 124 (2001) 5. — 23. EXCOFFIER L, LAVAL G, SCHNEIDER S, Evol Bioinform Online, 1 (2005) 47. — 24. NEI M, Molecular Evolutionary Genetics (Columbia University Press, New York, 1987). — 25. LEBART L, MORINEAU A, WARWICK K, Multivariate descriptive statistical analysis: Correspondence analysis and related techniques for large matrices (Wiley and Sons, New York, 1984). — 26. BANDELT HJ, FORSTER P, ROHL A, Mol Biol Evol, 16 (1999) 37. — 27. COOPER G, AMOS W, HOFFMAN D, RUBINSZTEIN DC, Hum Mol Genet, 5 (1996) 1759. — 28. DELFINER P, Linear estimation of non-stationary spatial phenomena. Advanced geostatistics in the mining industry (Reidel, Dordrecht, 1976). — 29. WILSON IJ, BALDING DJ, Genetics, 150 (1998) 499. — 30. WILSON JF, WEISS DA, RICHARDS M, THOMAS MG, BRADMAN N, GOLDSTEIN DB, Proc Natl Acad Sci USA, 98 (2001) 5078. — 31. ALVAREZ L, SANTOS C, MONTIEL R, CAEIRO B, BAALI A, DUGOUJON JM, ALUJA MP, Am J Hum Biol, 21 (2009) 407. — 32. ADAMS SM, BOSCH E, BALARESQUE PL, BALLEREAU SJ, LEE AC, ARROYO E, LOPEZ-PARRA AM, ALER M, GRIFO MSG, BRION M, CARRACEDO A, LAVINHA J, MARTINEZ-JARRETA B, QUINTANA-MURCI L,

PICORNELL A, RAMON M, SKORECKI K, BEHAR DM, CALAFELL F, JOBLING MA, Am J Hum Genet, 83 (2008) 725. — 33. GONÇALVES R, FREITAS A, BRANCO M, ROSA A, FERNANDES AT, ZHIVOTOVSKY LA, UNDERHILL PA, KIVISILD T, BREHM A, Ann Hum Genet, 69 (2005) 443. — 34. BATTAGLIA V, FORNARINO S, AL-ZAHERY N, OLIVIERI A, PALA M, MYRES NM, KING RJ, ROOTSI S, MARJANOVIC D, PRIMORAC D, HADZISELIMOVIC R, VIDOVIC S, DROBNIC K, DURMISHI N, TORRONI A, SANTACHIARA-BENERECETTI AS, UNDERHILL PA, SEMINO O, Eur J Hum Genet, 17 (2009) 820. — 35. KING R, Neolithic Migrations in the Near East and Aegean. In: Ancient Human Migrations: A Multidisciplinary Approach (The University of Utah Press, Salt Lake City, 2009). — 36. CRUCIANI F, SANTOLAMAZZA P, SHEN P, MACAULAY V, MORAL P, OLCKERS A, MODIANO D, HOLMES S, DESTRO-BISOL G, COIA V, WALLACE DC, OEFNER PJ, TORRONI A, CAVALLI-SFORZA LL, SCOZZARI R, UNDERHILL PA, Am J Hum Genet, 70 (2002) 1197. — 37. ZALLOUA PA, PLATT DE, EL-SIBAI M, KHALIFE J, MAKHOUL N, HABER M, XUE Y, IZAABEL H, BOSCH E, ADAMS SM, ARROYO E, LOPEZ-PARRA AM, ALER M, PICORNELL A, RAMON M, JOBLING MA, COMAS D, BERTRANPETIT J, WELLS RS, TYLER-SMITH C, Am J Hum Genet, 83 (2008) 633. — 38. CHIARONI J, KING RJ, MYRES NM, HENN BM, DUCOURNEAU A, MITCHELL MJ, BOETSCH G, SHEIKHA I, LIN AA, NIK-AHD M, AHMAD J, LATTANZI F, HERRERA RJ, IBRAHIM ME, BRODY A, SEMINO O, KIVISILD T, UNDERHILL PA, Eur J Hum Genet, 18 (2009) 348. — 39. CADENAS AM, ZHIVOTOVSKY LA, CAVALLI-SFORZA LL, UNDERHILL PA, HERRERA RJ, Eur J Hum Genet, 16 (2008) 374. — 40. TOFANELLI S, FERRI G, BULAYEVA K, CACIAGLI L, ONOFRI V, TAGLIOLI L, BULAYEV O, BOSCHI I, ALU M, BERTI A, RAPONE C, BEDUSCHI G, LUISELLI D, CADENAS AM, AWADELKARIM KD, MARIANI-COSTANTINI R, ELWALI NE, VERGINELLI F, PILLI E, HERRERA RJ, GUSMAO L, PAOLI G, CAPELLI C, Eur J Hum Genet, 17 (2009) 1520. — 41. EL-SIBAI M, PLATT DE, HABER M, XUE Y, YOUHANNA SC, WELLS RS, IZAABEL H, SANYOURA MF, HARMANANI H, BONAB MA, BEHBEHANI J, HASHWA F, TYLER-SMITH C, ZALLOUA PA, Ann Hum Genet, 73 (2009) 568. — 42. DI GAETANO C, CERUTTI N, CROBU F, ROBINO C, INTURRI S, GINO S, GUARRERA S, UNDERHILL PA, KING RJ, ROMANO V, CALI F, GASPARINI M, MATULLO G, SALERNO A, TORRE C, PIAZZA A, Eur J Hum Genet, 17 (2009) 91. — 43. SENGUPTA S, ZHIVOTOVSKY LA, KING R, MEHDI SQ, EDMONDS CA, CHOW CE, LIN AA, MITRA M, SIL SK, RAMESH A, USHA RANI MV, THAKUR CM, CAVALLI-SFORZA LL, MAJUMDER PP, UNDERHILL PA, Am J Hum Genet, 78 (2006) 202. — 44. SCHRACK B, ATHEY T, WILSON J, Cluster analysis of extended Y-STR haplotypes leads to discovery of a large and widespread sub-clade of Y Haplogroup J2 pathway (abstract 994) (Annual Meeting of the American Society of Human Genetics, New Orleans, Louisiana, 2006). — 45. MALASPINA P, TSOPANOMICHALOU M, DUMAN T, STEFAN M, SILVESTRI A, RINALDI B, GARCIA O, GIPARAKI M, PLATA E, KOZLOV AI, BARBUJANI G, VERNESI C, PAPOLA F, CIAVARELLA G, KOVATCHEV D, KERIMOVA MG, ANAGNOU N, GAVRILA L, VENEZIANO L, AKAR N, LOUTRADIS A, MICHALODIMITRAKIS EN, TERRENATO L, NOVELLETTO A, Ann Hum Genet, 65 (2001) 339. — 46. FLORES C, MACA-MEYER N, GONZALEZ AM, OEFNER PJ, SHEN P, PEREZ JA, ROJAS A, LARRUGA JM, UNDERHILL PA, Eur J Hum Genet, 12 (2004) 855. — 47. ALONSO S, FLORES C, CABRERA V, ALONSO A, MARTIN P, ALBARRAN C, IZAGIRRE N, DE LA RUA C, GARCIA O, Eur J Hum Genet, 13 (2005) 1293. — 48. CRUCIANI F, LA FRATTA R, SANTOLAMAZZA P, SELLITTO D, PASCONE R, MORAL P, WATSON E, GUIDA V, COLOMB EB, ZAHAROVA B, LAVINHA J, VONA G, AMAN R, CALI F, AKAR N, RICHARDS M, TORRONI A, NOVELLETTO A, SCOZZARI R, Am J Hum Genet, 74 (2004) 1014. — 49. CRUCIANI F, LA FRATTA R, TORRONI A, UNDERHILL PA, SCOZZARI R, Hum Mutat, 27 (2006) 831. — 50. THOMAS MG, SKORECKIAD K, BEN-AMID H, PARFITT T, BRADMAN N, GOLDSTEIN DB, Nature, 394 (1998) 138. — 51. NEBEL A, FILON D, WEISS DA, WEALE M, FAERMAN M, OPPENHEIM A, THOMAS MG, Hum Genet, 107 (2000) 630. — 52. TEMPLETON A, Population Genetics and Microevolutionary Theory (John Wiley & Sons, New Jersey, 2006). — 53. ACQUARO E, Los fenicios en el Mediterráneo Central en la época de Tarteso. In: Los enigmas de Tarteso (Ediciones Cátedra, Madrid, 1999). — 54. AJBAR MACHMUÂ, Crónica anónima del siglo XI. Colección de obras arábigas de historia y geografía (Real Academia de la Historia, Madrid, 1867). — 55. GUICHARD P, Al-Andalus: Estructura antropológica de una sociedad islámica en Occidente (Barral Editores, Barcelona, 1976). — 56. CARO-BAROJA J, Los Moriscos del reino de Granada (Alianza Editorial, Madrid, 2003). — 57. GARCIA IGLESIAS L, Los judíos en la España Antigua (Ediciones Cristiandad, Madrid, 1978). — 58. CARO BAROJA J, Los judíos en la España moderna y contemporánea (Istmo, Madrid 1962). — 59. BEHAR DM, VILLEMS R, SOODYALL H, BLUE-SMITH J, PEREIRA L, METSPALU E, SCOZZARI R, MAKKAN H, TZUR S, COMAS D, BERTRANPETIT J, QUINTANA-MURCI L, TYLER-SMITH C, WELLS RS, ROSSET S, Am J Hum Genet, 82 (2008) 1130.

*R. Calderón*

*Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense, Ciudad Universitaria, 28040 Madrid, Spain*
*e-mail: rcalfer@bio.ucm.es*

## ISTRAŽIVANJE NASELJAVANJA IBERIJSKOG POLUOTOKA IZ PERSPEKTIVE DVIJE ANDALUZIJSKE SUBPOPULACIJE: ISTRAŽIVANJE TEMELJENO NA J I E HAPLOGRUPAMA Y KROMOSOMA

## S A Ž E T A K

Cilj ovog istraživanja je viskorezolucijska analiza haplogrupa J i E Y kromosoma među andaluzijskim stanovnicima kako bi se rekonstruirale neolitičke, pretpovijesne i povijesne migracije u mediteranskoj regiji. Provedeno je genotipiziranje dva uzorka s Grenade (n=250 muškaraca) i s Huelve (n=167 muškaraca) (Španjolska) na binarnim i mikrosatelitnim markerima na Y kromosomu te su rezultati uspoređeni s ostalnim mediteranskim populacijama. Ova dva uzorka pokazuju genetičku različitost koja se može povezati s različitim evolucijskim procesima. Migracije prema Andaluziji vjerojatno su započele na Arapskom poluotoku, Levantu, Balkanu i sjevernoj Africi, predominantno u pretpovjesnim i povjesnim vremenima. Moreplovstvo je značajno doprinjelo nedavnom priljevu gena na Iberijski poluotok. Ovaj pregled naglašava složenost procesa mediteranskih migracija te pokazuje utjecaj različitih izvora populacija na genetičku kompoziciju španjolskih populacija. Glavne imigracije na Iberijski poluotok se vrlo vjerojatno nisu odvijale u međufazama, a ako jesu, onda ih je bilo samo nekoliko.