Heredity (2011) 106, 37–45

www.nature.com/hdy

ORIGINAL ARTICLE

Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge

O García[1,4], R Fregel[2,4], JM Larruga[2], V Álvarez[3], I Yurrebaso[1], VM Cabrera[2] and AM González[2]
[1]*Basque Country Forensic Genetics Laboratory, Erandio, Bizkaia, Spain;* [2]*Área de Genética, Departamento de Parasitología, Ecología y Genética, Facultad de Biología, Universidad de La Laguna (ULL), La Laguna, Tenerife, Spain and* [3]*Unidad de Genética, Hospital Universitario Central de Asturias, Oviedo, Asturias, Spain*

It has been proposed that the distribution patterns and coalescence ages found in Europeans for mitochondrial DNA (mtDNA) haplogroups V, H1 and H3 are the result of a post-glacial expansion from a Franco-Cantabrian refuge that recolonized central and northern areas. In contrast, in this refined mtDNA study of the Cantabrian Cornice that contributes 413 partial and 9 complete new mtDNA sequences, including a large Basque sample and a sample of Asturians, no experimental evidence was found to support the human refuge-expansion theory. In fact, all measures of gene diversity point to the Cantabrian Cornice in general and the Basques in particular, as less polymorphic for V, H1 and H3 than other southern regions in Iberia or in Central Europe. Genetic distances show the Cantabrian Cornice is a very heterogeneous region with significant local differences. The analysis of several minor subhaplogroups, based on complete sequences, also suggests different focal expansions over a local and peninsular range that did not affect continental Europe. Furthermore, all detected clinal trends show stronger longitudinal than latitudinal profiles. In Northern Iberia, it seems that the highest diversity values for some haplogroups with Mesolithic coalescence ages are centred on the Mediterranean side, including Catalonia and South-eastern France.
Heredity (2011) **106,** 37–45; doi:10.1038/hdy.2010.47; published online 21 April 2010

Keywords: mtDNA haplogroups; humans; Franco-Cantabrian refuge theory

Introduction

The distribution pattern and coalescence age detected for mitochondrial DNA (mtDNA) haplogroup V in European human populations has been interpreted as a molecular signal of a post-glacial expansion from a Franco-Cantabrian refuge that recolonized the almost depopulated central and northern areas of Europe, reflecting the Magdalenian colonization (Torroni *et al*., 1998, 2001). Although a mtDNA analysis in ancient Basque populations questioned that interpretation (Izagirre and de la Rua, 1999), the posterior molecular dissection of mtDNA haplogroup H, the dominant lineage in Europe, identified two subhaplogroups (namely H1 and H3) that displayed frequency distributions and coalescence ages very similar to those previously reported for haplogroup V (Achilli *et al*., 2004). As these patterns were confirmed by several independent studies (Loogvali *et al*., 2004; Pereira *et al*., 2005; Alvarez-Iglesias *et al*., 2009), this human refuge theory to explain the post-glacial resettlement of Europe has gained almost unanimous support. However, in our opinion, the issue is not yet satisfactorily settled. First, it seems that Central Europe was never completely depopulated even in Pleniglacial times (Terberger and Street, 2002) and that, before the Magdalenian colonization, contacts between Eastern and Western Europe seem to have existed as deduced from the affinities between the French Badegoulian and Danubian Kasovian cultures (Svoboda, 2007). At a genetic level, it is still necessary to confirm the distribution patterns and diversity levels of those lineages involved in little or not studied areas of the Cantabrian Cornice, to contrast these patterns with other maternal lineages that present similar coalescence ages and to assess whether further lineage subdivisions display the same correlated distributions. These are the goals of this study.

Correspondence: Dr R Fregel, Departamento de Genética, Facultad de Biología, Universidad de La Laguna, La Laguna, Santa Cruz de Tenerife 38270, Spain.
E-mail: rfregel@gmail.com
[4]*These authors contributed equally to this work.*
Received 1 December 2009; revised 23 February 2010; accepted 18 March 2010; published online 21 April 2010

Materials and methods

Sample sequencing and restriction fragment length polymorphism analysis

A total of 413 maternally unrelated individuals from different European areas were analysed for the hypervariable segments I (HVSI) and II (HVSII) of the mitochondrial control region (Gonzalez *et al*., 2006). The resulting haplotypes were provisionally sorted into main haplogroups following the updated nomenclature (van Oven and Kayser, 2009). To confirm this HVS-based haplogroup classification, all individuals assigned to a specific haplogroup were additionally tested by restriction

analysis of the diagnostic coding-region mutations (Supplementary Table S1), proposed to unambiguously classify sequences into haplogroups (Richards *et al*., 2000). In addition, 236 individuals belonging to the H haplogroup were further assorted into 1 of 19 different H subgroups (Supplementary Table S2), using diagnostic restriction fragment length polymorphisms (Ennafaa *et al*., 2009). We compiled 5436 sequences from the literature for total haplogroup statistical analysis (Supplementary Table S3), 1369 for H subgroup general comparisons (Supplementary Table S4), 1628 for the specific subgroups H1 and H3 analysis (Supplementary Table S5) and 724, 408, 189, 292 and 134, belonging to haplogroups K, T2b, W, V and HV0, respectively, for diversity analysis (Supplementary Tables S6 and S7). For our samples, the geographical origin of the oldest known maternal ancestor (three generations for the most) was considered as the sample source. Written informed consent was obtained from all individuals. This study was approved by the research ethics committee of the University of La Laguna.

For complete sequencing, the mitochondrial genome was amplified in 32 overlapping fragments with the primers and PCR conditions described previously (Maca-Meyer *et al*., 2001). The same primers were used to directly sequence the fragments on an ABI 3100 Analyser using Big-Dye Terminator chemistry (Applied Biosystems, Foster City, CA, USA). Sequence data were assembled and compared using the SeqScape software (Applied Biosystems), and all chromatograms were inspected independently by two researchers.

Data analysis

We have considered all the northern Spanish regions that face the Bay of Biscay as belonging to the Cantabrian Cornice, which from west to east are: Galicia, Asturias, Cantabria and the Basque Country. In addition, Northern Spain includes Aragon and Catalonia (Figure 1). For genetic comparisons with published data, only HVSI positions from 16 024 to 16 365 were taken into account. To use adequate sample sizes, small samples from the same area were pooled whenever they did not show statistical heterogeneity. AMOVA and F_{ST} distances were calculated on the basis of haplogroup and haplotype frequencies using ARLEQUIN 2.0 (Schneider *et al*., 2000). Multidimensional scaling plots were obtained from F_{ST} distances using SPSS version 17.0 (SPSS Inc., Chicago, IL, USA). Gene diversity was measured as *H*-values (Nei, 1987) and π-values (Tajima, 1983). Measures of haplotype richness were corrected for variation in sample size using the rarefaction method (Hurlbert, 1971) as implemented in ADZE 1.0 (Szpiech *et al*., 2008). Phylogenetic relationships and ρ-values among complete mtDNA sequences were established using the reduced median network algorithm (Bandelt *et al*., 1999) as implemented in the program Network 4.5.1.0 (Fluxus Engineering, Clare, UK; http://www.fluxus-engineering.com). Age estimations were calculated using the most accurate mutation rates proposed recently (Soares *et al*., 2009) for the entire molecule, correcting for selection (1 mutation per 2585 years), and only for synonymous positions (1 mutation per 7884 years) and compared with 1 mutation per 5138 years, the previously most used coding-region rate (Mishmar *et al*., 2003).

Figure 1 Map showing the Cornice region, comprising Galicia (GAL), Asturias (AST), Cantabria (CAN) and the Basque Country (BAS), and the other northern Spanish areas, Aragon (ARA) and Catalonia (CAT).

Results

Basque and Asturian mtDNA haplogroup and haplotype profiles

Supplementary Table S1 lists our 413 new sampled sequences according to their haplogroup status. A total of 295 different haplotypes were found, representing 71.4% of the total analysed sample. However, the proportion of different haplotypes in North-west Iberia (90%) is higher than in the adjacent Basque Country (64%). In spite of the severe isolation proposed to explain the genetic characteristics of the Basques, it seems that, as occurred for other populations of the Cantabrian Cornice (Maca-Meyer *et al*., 2003; Alvarez-Iglesias *et al*., 2009), they have also received external maternal influences. This fact is attested by the presence of an East-African-specific L3h1a2a lineage (confirmed by transition 156 46) and single representatives of the North-African clades M1 and U6 in the Basques (Supplementary Table S1). Although it is difficult to date the arrival of these lineages, the presence of an M1 haplotype in a Basque cemetery from the sixth to the seventh century AD, before the Arab occupation of Iberia (Alzualde *et al*., 2006), is in support of an old although small North-African influence on Northern Iberia. The rest of the haplotypes sampled are of clear Eurasian origin, and related to both Palaeolithic and Neolithic introductions.

One Neolithic haplogroup that shows differences in the Cantabrian Cornice is J, for which Asturians have the highest frequencies (9%), resembling the Galicians (8.6%) and differing from our Basque samples that, with the exception of the miscellaneous group (9.5%), show a mean frequency of 3%. These Basque J frequencies are in agreement with those found in previous studies (Bertranpetit *et al*., 1995; Corte-Real *et al*., 1996), but contrast with those (14.6%) obtained by Alfonso-Sanchez *et al*., 2008 for a mixed Vizcaya/Guipuzcoa sample. Similar high frequencies for haplogroup J were found in prehistoric (16%) and historic (16%) Basque remains (Izagirre and de la Rua, 1999; Alzualde *et al*., 2005). It has been stated that haplogroup J is one of the lineages that

mark the relative demic influence of the Neolithic in Europe (Richards *et al*., 2000). Thus, in contrast to the proposed lack of Neolithic intrusion in the Basque population (Bertranpetit *et al*., 1995), these data could be considered as evidence of Neolithic or post-Neolithic influences in them. However, they could also been explained supposing isolation and drift effects.

Another peculiarity that distinguishes the Northern Basque provinces of Guipuzcoa and Vizcaya from the rest is their high frequency of the Palaeolithic European U5 lineage (Supplementary Table S3). In fact, the U5b subgroup (characterized by transitions 16 192 and 16 270 in HVSI and by transition 150 in HVSII) shows the highest Iberian peak (15.4%) in Guipuzcoans. In contrast, U5b is rare among French Basques. This discrepancy and the prevalence of haplogroup J in the French Basques significantly differentiate them from the Spanish Basques (Richard *et al*., 2007).

Recently, dissection of the rare haplogroup U8a by complete mtDNA sequencing (Gonzalez *et al*., 2006) detected a great diversity among the few Basque lineages found. Phylogenetic coalescence ages showed that these lineages also had Palaeolithic roots in this region. Although always in low frequencies, U8a lineages were found from the Basque Country to the Volga-Ural region (Gonzalez *et al*., 2006), signalling a wide geographic spread since Palaeolithic times. In this wide distribution, another peak in frequency and diversity was detected in South-eastern France (Dubut *et al*., 2004). In this screening, haplogroup U8a had a mean frequency of 1.3% in Basques, not being detected to the west of the Basque Country but scattered over Catalonia, in the east of Spain (Supplementary Table S3). Thus, the suggestion that the southernmost dispersion of this haplogroup could be centred on South-east France instead of the Basque Country (Dubut *et al*., 2004; Alfonso-Sanchez *et al*., 2008) cannot be rejected.

Additional striking features that also differentiate the Basque provinces between them are the great diversity and high frequency for subhaplogroup X2 (Supplementary Table S3), and also the high frequency of haplogroup T (13.5%) in Vizcaya compared with those found in Guipuzcoa and Alava (~4%).

Other haplogroups that also show differences in the Cantabrian Cornice are the para-haplogroup HV* and haplogroups H and V, which will be discussed later.

mtDNA haplogroup distributions in the Cantabrian Cornice and surrounding areas

Haplogroup frequencies and sample sizes for all the populations analysed in this study are listed in Supplementary Table S3. AMOVA analysis showed that although the bulk of the total variance (98.9%) is due to differences among individuals within populations, geographic partitions within (0.9%) and among areas (0.2%) are also statistically significant at the 0.001 and 0.01 levels, respectively. It is worth mentioning that this microgeographic differentiation is most prominent in the Cantabrian Cornice.

Intra-regional differences were also detected in France (Dubut *et al*., 2004; Richard *et al*., 2007). Frequencies for haplogroups H, K, V, T2b and J* showed sharp differences in this area. Most of these disparities may be explained by the outlying location of the French

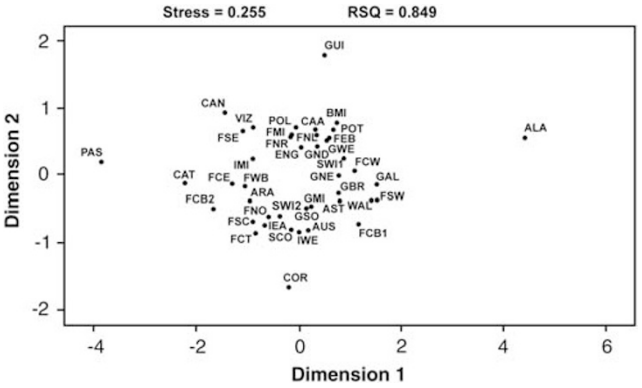

Figure 2 MDS plot based on linearized F_{ST} distances, comparing all the studied populations.

Basques, who are also quite different from the Spanish Basques (Supplementary Table S3) (Richard *et al*., 2007). Similarly, in the British Isles, Cornwall is by far the most discrepant sample; there are haplogroups such as K, T2b or V that showed the lowest frequencies in COR and the highest in Irish samples, whereas I and J* are more conspicuous in COR than in Ireland. Finally, within Central Europe, notable differences were also observed. For instance, J/J1c and J2a are fairly less represented in GWE than in northern samples. Differences are detectable even between southern samples, such as V that has a lower frequency in Austria compared with SWI1. However, when these differences are quantified by pairwise F_{ST} genetic distances (Supplementary Table S8), Central Europe, France (except the Basque sample) and the British Isles (without Cornwall and Wales), can be considered rather homogeneous areas.

In contrast, Northern Iberia appears to be microgeographically differentiated. Excluding the highly divergent Pasiego isolate (Maca-Meyer *et al*., 2003), there are also significant differences between the Basque provinces or between Catalonia at the north-eastern edge and Galicia-Asturias in the northwest. In fact, when these results were represented in an multidimensional scaling plot (Figure 2), the Pasiegos and Spanish Basques from Guipuzcoa and Alava were the most outstanding outliers, also followed by samples from Catalonia and Galicia, the French Basque sample and the British samples from Cornwall and Wales. Thus, contrasting with other European regions, the Cantabrian Cornice appears more as a sequence of small isolated populations, owing to its steep relief, than as a homogeneous population with a potential to expand widely, which contrasts with the idea that this area was an important source of post-glacial expansion. However, from an mtDNA perspective, the human Franco-Cantabrian glacial refugee hypothesis has been supported by the phylogeography and relative genetic diversity of haplogroup V and subgroups H1 and H3. Therefore, we have undertaken particularized analyses for these lineages.

Phylogeography and genetic diversities of HV* and V haplogroups in the Cantabrian Cornice and other European areas

The Palaeolithic para-haplogroup HV* is the ancestor of H and V haplogroups. It is moderately frequent in the

Cantabrian Cornice, but absent or rare in the French Basques (Richard *et al*., 2007). Its HV0 branch has the highest frequency in Northern Africa (Supplementary Table S9). However, there it presents the lowest gene and nucleotide diversities, excluding this area as a possible focus of expansion. In Europe, the highest HV0 frequencies are found in Northern Iberia and France, but its highest diversity is observed in NCE. In relation to the Iberian Peninsula, it shows its lowest diversities precisely in the Cornice area (IPNW), which is at the same level as in Northern Africa (Supplementary Table S9).

The first time haplogroup V was proposed as a sign of post-glacial human recolonization of Northern Europe from a Franco-Cantabrian refuge (Torroni *et al*., 1998), V frequencies in Basques (20%) and Catalans (24%) were found to be surprisingly high. However, this should now be considered as due to sampling errors because when sample sizes were increased in posterior analyses, V frequencies in the Basque Country dropped to 12.4% (Torroni *et al*., 2001) and 10.2% (Maca-Meyer *et al*., 2003). In this study, haplogroup V frequencies in the Cornice are at their peak in Cantabria (19%), dropping westwards to 5.6% in Asturias and to 3.8% in Galicia. In the Basque Country, haplogroup V frequencies ranged from 11.7% in Guipuzcoa to 5.9% in the Alava province. Finally, in a recent survey (Alvarez-Iglesias *et al*., 2009), V frequencies for Catalonia were estimated at around only 3%. Diversity values for V are significantly higher in Southern Iberia than in the Cornice ($P<0.05$). Excluding Scandinavia, the lowest diversities are found in Northern Africa and the Iberian northeast.

The post-glacial refuge expansion of V from a Franco-Cantabrian refugee hypothesis did not receive unanimous acceptation. It was first questioned on the basis of a lack of V representatives in ancient Basque samples (Izagirre and de la Rua, 1999) and its later presence in a historical sample from Alava (Alzualde *et al*., 2005), and second, for a lack of any directional gene-flow process of V along the proposed north-west European transects (Simoni *et al*., 2000). It was suggested that perhaps the Cantabrian area was a more probable expansion centre than the Basque Country (Maca-Meyer *et al*., 2003). However, the lower diversity values found here for this putative area compared with Southern Iberia also weaken this alternative. Furthermore, a recent mtDNA study of French populations included in the hypothetical last glacial maximum refuge stand out by their shortage of V lineages (Dubut *et al*., 2004).

Updated phylogeny of haplogroup HV4

Recently, a dissection of para-haplogroup HV* in eastern European populations has been performed using complete sequences (Malyarchuk *et al*., 2008). A new clade, named HV4, characterized by transition 7094, was identified. In this study, we detected four subjects who harboured the HV4 diagnostic transition 7094. Complete sequencing of these lineages and the addition of another seven complete or nearly complete HV4 sequences, published by other authors, allowed us to construct a phylogenetic tree for this little-studied clade (Figure 3). All our samples from the Cantabrian Cornice fall into a branch defined by transition 13 680, provisionally named HV4a1, which also includes one Italian and one European lineage. As the three samples from the Cantabrian Cornice also shared transition 16 291, we grouped them into a more specific subbranch named HV4a1a. Another new branch, here named HV4a2, is defined by transition 7085 and comprises two geographically

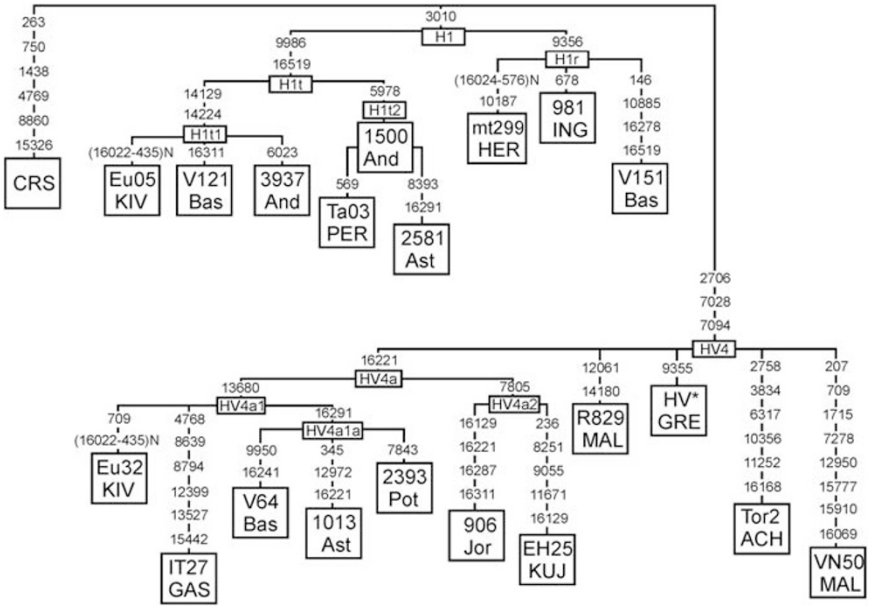

Figure 3 Phylogenetic tree based on complete H1t, H1r and HV4 sequence numbers along links refer to nucleotide positions. Regions not analysed are in parentheses. GenBank accession numbers of the subjects retrieved from the literature are: Eu05KIV (DQ112766; Spain) and Eu32KIV (DQ112831; Europe) from Kivisild *et al*. (2006); Ta03PER (EF177427; Portugal) from Pereira *et al*. (2007); R829MAL (EU545447; Russia) and VN50MAL (EF222234; Slav) from Malyarchuk *et al*. (2008); HV*GRE (EF417833) deposited by Greenspan, B. from Family Tree DNA; EH25KUJ (EU935457; Egypt) from Kujanova *et al*., 2009; Tor2ACH (AY738941; Italy) from Achilli *et al*. (2004); IT27GAS (EF660939; Italy) from Gasparre *et al*. (2007); mt299HER (EF657452; Europe) from Herrnstadt *et al*. (2002); 981ING (AF346981; France) from Ingman *et al*. (2000). Population origins of our samples are: V121 and V151, Basques from Guipuzcoa; V64, Basques from Vizcaya; 3937 and 1500 from Andalusia; 2581 and 1013, from Asturias; 2393, from Potes (Cantabria); and 906 from Jordan.

closed lineages, our Jordanian 906 and the Egyptian EH25 (Kujanova *et al*., 2009). In a recent global mtDNA phylogenetic tree, based on published complete sequences (van Oven and Kayser, 2009), a branch named HV4a was proposed using transition 16 221 as the diagnostic position. From our tree, it can be deduced that this position could indeed be ancestral to the HV4a1 and HV4a2 branches, although it suffered two retromutations, one in each branch. Therefore, although its use as diagnostic position seems to be problematic, by parsimony, we have kept using transition 16 221 to define the HV4a branch. Clearly, ages (Supplementary Table S10) for the whole HV4 clade (15.9 ± 3.5 ky or 13.2 ± 1.0 ky) and for the HV4a1 branch (10.3 ± 3.2 ky or 9.7 ± 0.2 ky), where our samples from the Cantabrian Cornice are included, and the Valencian HV4a2 branch (7.7 ± 4.4 ky or 9.9 ± 2.1 ky), are compatible with multiple post-glacial reexpansions with different geographic origins, including the Near East.

Phylogeography of H subhaplogroups in the Cantabrian Cornice and other European areas

In accordance with previous screenings, haplogroup H is predominant (Supplementary Table S2) in the Cantabrian Cornice. The subdivision of this clade into subhaplogroups uncovers significant differences even between the geographically close Basque provinces. For instance, the recently characterized H2a5 autochthonous Basque sublineage (Alvarez-Iglesias *et al*., 2009) has all its representatives in Guipuzcoa (17%) but one was found in the Vizcaya province. Frequencies for H3 are also higher in the former (19%) than in the latter (10.8%). On the other hand, the high frequency of H4 haplotypes in Vizcaya (8.1%) contrasts with their absence in Guipuzcoa. However, the bulk of the H lineages in all the samples have an H1* assignation (Supplementary Table S4). The most notable peculiarity of haplogroup H dissection in Asturians was the relatively high frequency (23%) of unclassified H* lineages that contrasts with the low Basque percentage (9.7%) and links Asturians to their Galician neighbours to the west (21%).

As detected before, frequencies for subgroups H1 and H3 are particularly high in the Cantabrian Cornice. However, sublineage H1a is only present in the Pasiegos, who also harbour the highest frequency for subgroup H4 (18.2%) and a surprising lack of H3 lineages. It has been previously suggested that the Pasiegos are a complex human isolate. On the basis of their high frequency for the North-African Y-chromosome E-M81 haplogroup and a relative abundance of other Y-chromosome and mtDNA haplogroups, of typical northern European ascription, dual gene-flow influences were assumed to affect this isolate (Maca-Meyer *et al*., 2003). In this study, again, the high frequency of H4 and the absence of H3 resembles the haplogroup H profile found for Moroccan Berbers (Ennafaa *et al*., 2009). On the other hand, the high frequency of H1a may be better explained by north-central European influence (Loogvali *et al*., 2004).

Another striking result is the high frequency found in Cantabrians for H6a (12.9%), which contrasts with its overall scarcity in the rest of the Cornice. However, as a single haplotype (16 362) is responsible for that frequency, this peak could be attributed to genetic drift and relative isolation effects.

Finally, the high frequency of H7 lineages, particularly concentrated in North-east Iberia and Southern France, suggests a local focus of expansion. However, as H7 is also comparatively frequent in Tunisia and the Near East (Roostalu *et al*., 2007; Ennafaa *et al*., 2009), a Mediterranean propagation of this clade seems more in accordance with its geographical pattern. As subgroup H1 and H3 patterns are in support of this post-glacial recolonization hypothesis, a more detailed analysis of these clades is presented in the following.

Phylogeography of H1 in the Cantabrian Cornice and other European areas

Frequencies for H1 and its subclades are reported in Supplementary Table S11 for the regions analysed. Although the highest frequency for the whole H1 was found in the Cantabrian Cornice as reported previously (Achilli *et al*., 2004; Loogvali *et al*., 2004; Pereira *et al*., 2005; Alvarez-Iglesias *et al*., 2009), diversity values were significantly higher in North-east Iberia in particular, and in the Mediterranean area in general. Furthermore, the allelic richness in the Cornice and in France is the lowest in all the studied range. In addition, the H1 frequency distribution pattern has to be taken with caution because it comprises other subclusters with very heterogeneous geographic patterns (Loogvali *et al*., 2004). Thus, H1a is most abundant in an area centred on Eastern Europe where the greatest diversity and allelic richness has been detected, being practically absent in Iberia (Supplementary Table S11). On the other hand, H1b is a rather scarce subclade scattered across Europe, the diversity and allelic richness of which peak in Southern Iberia but not the Cornice (Supplementary Table S11).

F_{ST} pairwise comparisons based on haplotype frequencies detected unexpected heterogeneity. France showed close affinities with only the nearby north-east Iberian sample. In addition, the Scandinavians seem to be very different from north-central Europeans, showing more affinities to Slavs. The rest of the H1 haplotypes (grouped as H1*) are most probably genealogically heterogeneous, including new subclades with different geographical patterns. However, it is worth mentioning that H1* showed a longitudinal cline ($r^2 = 0.690$; $P = 0.04$), with frequencies decreasing to the east as previously detected for the whole haplogroup H. Furthermore, there was also a positive correlation between the spatial diversities of H1 and H1* ($r = 0.801$; $P = 0.010$). On the other hand, it seems that H1 as a whole has a higher diversity in the Near East than in Iberia (Ennafaa *et al*., 2009). In short, we find no well-founded reasons to confirm that the H1 distribution in Europe reflects a human expansion centred on the Franco-Cantabrian area.

Updated phylogeny of haplogroups H1r and H1t

Screening by restriction fragment length polymorphism, and subsequent complete sequencing for some unclassified H1 lineages, unveiled the presence of rare subgroups such as H1r and H1t in the Basque Country (Figure 3). H1r is defined by transition 9356 (van Oven and Kayser, 2009) and seems to have a European range of distribution. In turn, H1t, defined by transition 9986, can be split into two branches: H1t1 characterized by transitions 14 129 and 14 224, and H1t2 characterized by transition 5978 (Figure 3). Contrarily to

H1r, the subclade H1t seems to be confined to the Iberian Peninsula as all the lineages analysed up to now are of Spanish or Portuguese ancestry. Curiously, estimated coalescence ages (Supplementary Table S10) for H1t (9.4 ± 4.6 ky or 5.8 ± 0.5 ky) are older than or similar to H1r (5.1 ± 3.0 ky or 5.2 ± 0.0 ky). Subclade H1t was also detected once in Asturians but it belongs to the H1t2 branch.

Phylogeography of H3 in the Cantabrian Cornice and other European areas

The highest European frequencies for H3 in Europe are found in the Cantabrian Cornice and in Southern Iberia (Supplementary Table S12), but again, the Cornice and France are the regions showing the lowest diversity values and, like the H1a subclade, the highest diversity and allelic richness for H3 in Europe are found in north-eastern and north-central regions. A significant longitudinal cline, similar to that found for H1* is detected for H3 ($r = 0.864$; $P = 0.003$) and, again, with frequencies decreasing eastwards. In addition, the Iberian Peninsula populations show strong haplotype differentiation when compared with North-central Europe and Scandinavia. Although, there is an H3 frequency peak centred on Iberia, comparative diversity values are against the suggestion that this area was the expansion centre for the European colonization of this clade.

Phylogeography of haplogroups (K, T2, W) with a probable Mesolithic expansion in Europe

Several haplogroups, including K, T2b and W, have been considered to expand in Europe during the late Upper Palaeolithic–Mesolithic (Richards *et al*., 2000). To compare their distribution patterns with those obtained for V, H1 and H3, their relative frequencies and diversity values in the Cantabrian Cornice and other European areas were determined (Supplementary Table S13). Although K frequencies in France and the British Isles are the highest, gene and nucleotide diversities peak in North-east Iberia and in Scandinavia. However, the lowest frequency and diversities were found in the Slavs; indeed there is a significant longitudinal cline for the K gene diversity in Europe ($r = 0.745$; $P = 0.022$), with increasing values to the west. T2b and W show strikingly similar patterns. Both have their highest diversity values in Northern Iberia, including the Cantabrian Cornice, and in the Mediterranean area. In contrast, their diversities decreased significantly in neighbouring France ($P < 0.001$ in both cases) and in the British Isles ($P < 0.001$ in both cases) pointing more to a Mediterranean-centred than a latitudinal Atlantic expansion.

Discussion

Population analysis

Our population mtDNA analysis of the Northern Iberian Peninsula detected a significant genetic microdifferentiation affecting mainly the Basque Country. Distances between Basque samples are greater than between them and samples from other regions, pointing to strong local isolation and limited gene-flow interchange between them and the surrounding regions. In contrast, Galicia, Asturias and the Cantabrian village of Potes, on the Asturian border, to the west, and Catalonia and Aragon to the east of the Basque Country form rather homogeneous regions that are well differentiated between them.

Cantabria shows more affinities with the geographically closer Vizcayan population than with the other Basque samples. Although at a regional level this pattern is congruent with an isolation by distance model with limited maternal gene flow, at a local level, mainly in the Basque Country, it is better explained as result of a pronounced isolation and very limited focal expansions. In addition, the fact that the haplogroups show different distributions point to the overlapping of several radiating waves, spatially and through time.

When these analyses were extended to the rest of Europe, the general picture was significantly more homogeneous than that of Northern Iberia. We have not found any special mtDNA link between Basques and the British Isles, in agreement with published mtDNA (McEvoy *et al*., 2004) and Y-chromosome (Alonso *et al*., 2005) studies, and contrary to previous suggestions about a particular link between Basque and Celtic populations (Wilson *et al*., 2001).

In general, our results are congruent with the apparent lack of mtDNA structure found for continental Europe compared with the more differentiated Mediterranean peninsulas (Simoni *et al*., 2000). These results are also in accordance with the different patterns found for the Y-chromosome haplogroup distribution in continental Europe compared with the Mediterranean peninsulas, with smooth clines and gradients apparent in continental areas (Rosser *et al*., 2000; Semino *et al*., 2000) contrasting to the focal expansions detected in Iberia (Flores *et al*., 2004), Italy and Greece (Di Giacomo *et al*., 2003).

Recently, genome-wide studies, conducted to assess the level of European population stratification, have shown a close correspondence between genetic and geographic distances (Novembre *et al*., 2008). At the population level, the most prominent patterns uncovered were a consistent distinction between Southern Mediterranean and northern continental Europeans (Seldin *et al*., 2006) and a clear separation of northern from south-eastern Europeans (Bauchet *et al*., 2007). This is in line with mtDNA results and consistent with the clines observed using classical markers (Menozzi *et al*., 1978) and some mtDNA (Richards *et al*., 1996) and Y-chromosome haplogroups (Rosser *et al*., 2000; Semino *et al*., 2000). These analyses also clearly separate Iberians including Basques as distinct from other Europeans but, in accordance with our results, they did not find specific affinities between Basques and the English and Irish, who cluster clearly with the continental Germans and Poles (Bauchet *et al*., 2007).

From these results, it can be deduced that although modern Iberians share a common mtDNA genetic background with other Europeans, they have evolved, at least in the north, as discrete genetic clusters with limited local expansions and not as a genetic continuum extensible to the rest of Europe.

Phylogeographic analysis

In addition to the broadly distributed mtDNA haplogroups, the Cantabrian Cornice and the Basque Country in particular are characterized by the presence of specific subclades with different coalescence ages that testify to

the existence of several localized expansions. However, the lack of correlation among them is evidence against the existence of a major migration wave with continental repercussions.

The high frequency and variability found for haplogroup U5b and the wide differentiation detected among U8a Basque lineages (Gonzalez *et al*., 2006) are signals that, at least, the Basque population preserves maternal Palaeolithic lineages in its present-day maternal pool. It is also evident that the Basques underwent local expansions or participated in Iberian Peninsula range expansions at different Neolithic and post-Neolithic periods.

In a previous study, a Basque-specific H2a5 subclade was described (Alvarez-Iglesias *et al*., 2009). This clade seems to have an overwhelming implantation in the province of Guipuzcoa showing a very recent expansion age (1.3 ± 0.9 ky). A Y-chromosome counterpart of this type of very localized lineages could be the R1-M65 subclade, detected in a small Basque sample (Bosch *et al*., 2001), which could not be confirmed in a larger screening (Alonso *et al*., 2005).

Possible Neolithic radiations within Cantabria and Iberia have also been detected, as attested by the coalescence age of the Cantabrian H1r lineage (5.2 ± 0.0) and for the ages of subgroup H1t (5.8 ± 0.5) and the HV4a1a branch (6.5 ± 1.3), both of a supposed Iberian range. Interestingly, the latter lineage belongs to the phylogenetically older HV4a1 European clade that may have expanded around 9.7 ± 0.2 ky. The Y-chromosome counterparts for these maternal expansions in Iberia could be the R-M153 and the R-SRY2627 subclades (Hurles *et al*., 1999; Bosch *et al*., 2001). Both were recently detected in the Spanish Pyrenees and updated ages of 8.4 ± 2.7 ky for the former and 7.4 ± 1.5 ky for the latter were estimated (Lopez-Parra *et al*., 2009).

Contrary to this, for those lineages for which a post-glacial European recolonization from a Franco-Cantabrian refuge was proposed, we have not found any evidence of such expansion in the Cantabrian Cornice. For instance, diversity values for haplogroup V are significantly higher in Southern Iberia than in the Cornice and, confirming previous results (Simoni *et al*., 2000), we have not found significant latitudinal clines for this haplogroup. Furthermore, focussing on haplotype sharing, the Cornice is the most divergent area, despite North-east Iberia being the closest region to the rest of Europe. This could mean that the proposed expansion had a French instead of a Cantabrian origin, but the recent mtDNA regional study on French samples weakens this possibility (Dubut *et al*., 2004).

As for the H1 and H3 subhaplogroups, the strong north-west–south-east European cline detected for the whole H haplogroup in early studies has been criticised as a chimera in the light of the complex structure revealed by the dissection of this haplogroup into subclades (Pereira *et al*., 2005). However, the same should be applied to the proposed H1 dispersion, as the H1a and H1b subbranches present very different distributions in Europe (Loogvali *et al*., 2004). On the other hand, the H1a subbranch is practically absent from the Iberian Peninsula and the highest diversity found in this region for the H1b subbranch is in Southern Iberia instead of in the Cornice. Furthermore, any clinal evidence for the residual paragroup H1* point to a north-west to south-east trend not to a south-west to north trend. Regarding the H3 subhaplogroup, in this and in recent studies (Alvarez-Iglesias *et al*., 2009) the Cantabrian Cornice and France have been independently confirmed to show the highest frequencies in Europe for it, but precisely these regions show its lowest diversity values. As occurs for H1a, the H3 highest diversity is in North-central Europe. The Y-chromosome I-M253 subclade has been proposed as another lineage signalling a northward expansion from the Iberian Peninsula-Southern France refuge after the LGM (Rootsi *et al*., 2004). However, Y-chromosome analysis of Iberia pointed to the I-M26 subclade as the most prevalent there (Flores *et al*., 2004; Alonso *et al*., 2005; Lopez-Parra *et al*., 2009). It has been suggested that a region of Iberia-Southern France could be the origin of this subclade (Rootsi *et al*., 2004). The high frequency and diversity found for it in Pyrenean populations (Lopez-Parra *et al*., 2009) seems to confirm that hypothesis. However, in contrast to I-M253, I-M26 has an extremely low frequency in Scandinavia and most probably had a minor role in the recolonization of this region (Rootsi *et al*., 2004).

Other mtDNA lineages, such as K, T2 and W, could mark the proposed Mesolithic expansion in Europe. However, haplogroup K has its highest diversity centred on North-eastern Iberia instead of the Cantabrian Cornice, and the highest diversities for haplogroups T2b and W are also centred on Northern Iberia, in this case including the Cornice. However, diversities are higher in the Mediterranean area than in Atlantic France and the British Isles.

Expansion ages for several mtDNA haplogroups coincide with a post-glacial period and it seems congruent to deduce that demic expansions occurred in that period, as the archaeological record suggests (Gamble *et al*., 2004). However, the model of a major human demic colonization of Central and Northern Europe from a Franco-Cantabrian refuge, is not supported experimentally, at least at a mtDNA level. Nevertheless, we have noticed that the highest diversity values for several mtDNA lineages are centred in a region comprising North-eastern Spain and South-east France instead of the Cantabrian Cornice. Archaeologically, this region also shows continuity of human settlement since Palaeolithic times and has been marked as an area of great Mesolithic, Neolithic and post-Neolithic importance. More studies at genomic level will be necessary to uncover the fine-grained mtDNA phylogeography of Europe.

Accession numbers

The nine new complete mitochondrial DNA sequences are registered under GenBank accession numbers GQ888723-31.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We acknowledge P Hernández Carpio for technical assistance. This research was supported by Grant no. BFU2006-04490 from the Ministerio de Ciencia y Tecnología to JML, and RF was supported by a Gobierno de Canarias predoctoral fellowship.

References

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R *et al.* (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* **75**: 910–918.

Alfonso-Sanchez MA, Cardoso S, Martinez-Bouzas C, Pena JA, Herrera RJ, Castro A *et al.* (2008). Mitochondrial DNA haplogroup diversity in Basques: a reassessment based on HVI and HVII polymorphisms. *Am J Hum Biol* **20**: 154–164.

Alonso S, Flores C, Cabrera V, Alonso A, Martin P, Albarran C *et al.* (2005). The place of the Basques in the European Y-chromosome diversity landscape. *Eur J Hum Genet* **13**: 1293–1302.

Alvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintans B, Zarrabeitia MT, Cusco I *et al.* (2009). New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS ONE* **4**: e5112.

Alzualde A, Izagirre N, Alonso S, Alonso A, Albarran C, Azkarate A *et al.* (2006). Insights into the 'isolation' of the Basques: mtDNA lineages from the historical site of Aldaieta (6th-7th centuries AD). *Am J Phys Anthropol* **130**: 394–404.

Alzualde A, Izagirre N, Alonso S, Alonso A, de la Rua C (2005). Temporal mitochondrial DNA variation in the Basque Country: influence of post-Neolithic events. *Ann Hum Genet* **69**(Pt 6): 665–679.

Bandelt HJ, Forster P, Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**: 37–48.

Bauchet M, McEvoy B, Pearson LN, Quillen EE, Sarkisian T, Hovhannesyan K *et al.* (2007). Measuring European population stratification with microarray genotype data. *Am J Hum Genet* **80**: 948–956.

Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995). Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* **59**(Pt 1): 63–81.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, 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.

Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A *et al.* (1996). Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**(Pt 4): 331–350.

Di Giacomo F, Luca F, Anagnou N, Ciavarella G, Corbo RM, Cresta M *et al.* (2003). Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Mol Phylogenet Evol* **28**: 387–395.

Dubut V, Chollet L, Murail P, Cartault F, Beraud-Colomb E, Serre M *et al.* (2004). mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur J Hum Genet* **12**: 293–300.

Ennafaa H, Cabrera VM, Abu-Amero KK, Gonzalez AM, Amor MB, Bouhaha R *et al.* (2009). Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet* **10**: 8.

Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, Perez JA *et al.* (2004). Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet* **12**: 855–863.

Gamble C, Davies W, Pettitt P, Richards M (2004). Climate change and evolving human diversity in Europe during the last glacial. *Philos Trans R Soc Lond B Biol Sci* **359**: 243–253, discussion 253–244.

Gasparre G, Porcelli AM, Bonora E, Pennisi LF, Toller M, Iommarini L *et al.* (2007). Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. *Proc Natl Acad Sci USA* **104**: 9001–9006.

Gonzalez AM, Garcia O, Larruga JM, Cabrera VM (2006). The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. *BMC Genomics* **7**: 124.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C *et al.* (2002). Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* **70**: 1152–1171.

Hurlbert SH (1971). The non-concept of species diversity: a critique and alternative parameters. *Ecology* **52**: 577–586.

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Perez-Lezaun A *et al.* (1999). Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* **65**: 1437–1448.

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708–713.

Izagirre N, de la Rua C (1999). An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major Paleolithic expansion from southwestern Europe. *Am J Hum Genet* **65**: 199–207.

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K *et al.* (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**: 373–387.

Kujanova M, Pereira L, Fernandes V, Pereira JB, Cerny V (2009). Near Eastern Neolithic genetic input in a small oasis of the Egyptian Western Desert. *Am J Phys Anthropol* **140**: 336–346.

Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E *et al.* (2004). Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* **21**: 2012–2021.

Lopez-Parra AM, Gusmao L, Tavares L, Baeza C, Amorim A, Mesa MS *et al.* (2009). In search of the pre- and post-Neolithic genetic substrates in Iberia: evidence from Y-chromosome in Pyrenean populations. *Ann Hum Genet* **73**: 42–53.

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* **2**: 13.

Maca-Meyer N, Sanchez-Velasco P, Flores C, Larruga JM, Gonzalez AM, Oterino A *et al.* (2003). Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet* **67**(Pt 4): 329–339.

Malyarchuk B, Grzybowski T, Derenko M, Perkova M, Vanecek T, Lazur J *et al.* (2008). Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Mol Biol Evol* **25**: 1651–1658.

McEvoy B, Richards M, Forster P, Bradley DG (2004). The Longue Durée of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. *Am J Hum Genet* **75**: 693–702.

Menozzi P, Piazza A, Cavalli-Sforza L (1978). Synthetic maps of human gene frequencies in Europeans. *Science* **201**: 786–792.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S *et al.* (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* **100**: 171–176.

Nei N (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A *et al.* (2008). Genes mirror geography within Europe. *Nature* **456**: 98–101.

Pereira L, Goncalves J, Franco-Duarte R, Silva J, Rocha T, Arnold C *et al.* (2007). No evidence for an mtDNA role in sperm motility: data from complete sequencing of asthenozoospermic males. *Mol Biol Evol* **24**: 868–874.

Pereira L, Richards M, Goios A, Alonso A, Albarran C, Garcia O *et al.* (2005). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* **15**: 19–24.

Richard C, Pennarun E, Kivisild T, Tambets K, Tolk HV, Metspalu E *et al.* (2007). An mtDNA perspective of French genetic variation. *Ann Hum Biol* **34**: 68–79.

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A *et al.* (1996). Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* **59**: 185–203.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V *et al.* (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**: 1251–1276.

Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, Reidla M *et al.* (2007). Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* **24**: 436–448.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M *et al.* (2004). Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* **75**: 128–137.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A *et al.* (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67**: 1526–1543.

Schneider S, Roessli D, Excoffier L (2000). *Arlequin ver. 2000: A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva: Geneva, Switzerland.

Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G *et al.* (2006). European population substructure: clustering of northern and southern populations. *PLoS Genet* **2**: e143.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE *et al.* (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* **290**: 1155–1159.

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

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A *et al.* (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* **84**: 740–759.

Svoboda JA (2007). The Gravettian of the middle Danube. *Paleo* **19**: 203–220.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**: 2498–2504.

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.

Terberger T, Street M (2002). Hiatus or continuity? New results for the question of pleniglacial settlement in Central Europe. *Antiquity* **76**: 691–698.

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D *et al.* (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 HJ, Macaulay V, Richards M, Cruciani F, Rengo C *et al.* (2001). A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69**: 844–852.

van Oven M, Kayser M (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**: E386–E394.

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* **98**: 5078–5083.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)