European Journal of Human Genetics (2005) 13, 1293–1302

www.nature.com/ejhg

ARTICLE

The place of the Basques in the European Y-chromosome diversity landscape

Santos Alonso*[,1], Carlos Flores[2], Vicente Cabrera[3], Antonio Alonso[4], Pablo Martín[4], Cristina Albarrán[4], Neskuts Izagirre[1], Concepción de la Rúa[1] and Oscar García[5]

[1]*Dpto. Genética, Antropología Física y Fisiología Animal, Fac. Ciencia y Tecnología, UPV/EHU, Barrio Sarriena s/n 48940, Leioa, Bizkaia, Spain;* [2]*Unidad de Investigación, Hospital Univ. N.S. de Candelaria, Tenerife 38010, Spain;* [3]*Dpto. de Genética, Fac. de Biología, Univ. de La Laguna, Tenerife 38271, Spain;* [4]*Instituto Nacional de Toxicología y Ciencias Forenses, Sección de Biología, Luis Cabrera 9, 28002 Madrid, Spain;* [5]*Área de Lab. Ertzaintza, Larrauri Mendotxe Bidea 18, 48950, Erandio, Bizkaia, Spain*

There is a trend to consider the gene pool of the Basques as a 'living fossil' of the earliest modern humans that colonized Europe. To investigate this assumption, we have typed 45 binary markers and five short tandem repeat loci of the Y chromosome in a set of 168 male Basques. Results on these combined haplotypes were analyzed in the context of matching data belonging to approximately 3000 individuals from over 20 European, Near East and North African populations, which were compiled from the literature. Our results place the low Y-chromosome diversity of Basques within the European diversity landscape. This low diversity seems to be the result of a lower effective population size maintained through generations. At least some lineages of Y chromosome in modern Basques originated and have been evolving since pre-Neolithic times. However, the strong genetic drift experienced by the Basques does not allow us to consider Basques either the only or the best representatives of the ancestral European gene pool. Contrary to previous suggestions, we do not observe any particular link between Basques and Celtic populations beyond that provided by the Paleolithic ancestry common to European populations, nor we find evidence supporting Basques as the focus of major population expansions.
European Journal of Human Genetics (2005) **13,** 1293–1302. doi:10.1038/sj.ejhg.5201482;
published online 10 August 2005

Keywords: Y chromosome; Basques; SNPs; STRs; compound-haplotypes; evolutionary history; M153

Introduction

An *ad hoc* mining of the historical record can lead to a spurious association of any finding in human population genetics with any historical episode that could potentially explain it.[1] In this context, the hypothesis of Basques representing an ancestral Paleolithic European population, already put forward on linguistic grounds in the 19th century, has been used as a recurrent explanation in a number of early population-genetic studies.[2–4] Currently, most linguists agree that the Basque language (Euskera) should be considered pre-Roman and pre-Indo-European, with no robust phylogenetic relationships. Consequently, its origins are lower-bounded in the second millennium BC. Linguistically, any other hypothesis positing a more ancient origin for Basque cannot be proven with the currently available scientific methodology.[5] However, the idea of a Basque genetic pool that shows little influence from both the Neolithic and later population flows, has spread through the literature as a circular argument that has led to use the Basque population as the representative gene pool of the first modern human settlers of Europe.[6–8] Thus, with the general aim to investigate the evolutionary

***Correspondence: Dr S Alonso, Dpto. Genética, Antropología Física y Fisiología Animal, Fac. Ciencia y Tecnología, UPV/EHU, Barrio Sarriena s/n 48940 Leioa, Bizkaia, Spain. Tel: +34 946013568; Fax: +34 946013500; E-mail: ggpalals@lg.ehu.es**
Received 4 April 2005; revised 14 June 2005; accepted 8 July 2005; published online 10 August 2005

history of the Basques we studied in this work a large Basque sample by means of high-resolution compound Y-chromosome haplotypes and analyzed them in the context of the available European, North African and Near East Y-chromosome data from the literature. The Y chromosome offers a series of advantages for these purposes. In particular, the nonrecombining nature of the Y chromosome facilitates the inference of compound haplotypes made up of slowly evolving single-nucleotide polymorphisms (SNPs) and more quickly evolving short tandem repeat (STR) loci, which offers the possibility to study the Y phylogeny at different resolutions and thus, at different time scales.

Materials and methods

Population samples

Y chromosomes from 168 unrelated Basque donors were used for the study: 72 from the province of Biscay, 74 from Gipuzkoa and 22 (Other Basques) from the Alava and Navarre provinces. Donors had at least four generations of ancestry in the Basque Country (recorded by Basque surnames), and within each of our three Basque samples (Biscay, Gipuzkoa and Others) all the grandparents of the donors were born in the same province. We also included in this study 459 non-Basque Iberians from diverse localities that had been partially genotyped previously[9] and which have been further genotyped in this work along with an additional set of 233 non-Basque Iberians, and 75 North-African Berbers. In addition, data from 39 European and Near-Eastern populations were compiled from the literature for comparative purposes (Supplementary Information 1). Among these, two other Basque samples were included: on the one hand, a Basque sample representing a general sample from Gipuzkoa[10,11] (E Bosch, pers. comm.), and on the other hand, the Basque sample of Brión *et al*,[12] whose precise geographical origin is unknown. Finally, some data (33 French from Normandy and 53 Georgians) correspond to unpublished work (JM Larruga, pers. comm.) and have been kindly submitted to us for comparative purposes.

Biallelic markers

A total of 45 binary markers have been analyzed. First, we genotyped nine genealogically basal markers in all individuals (SRY10831.1, YAP, M89, P2, M9, M201, M170, 12f2 and 92R7) and the remaining markers (M2, M12, M13, M20, M26, M34, M52, M65, M67, M70, M78, M81, M92, M107, M122, M123, M124, M148, M153, M163, M165, M166, M172, M173, M175, M178, M207, M224, M269, M342, M377, P15, P16, SRY10831.2, SRY2627 and Tat) were genotyped following the hierarchy of the genealogy.[13] Markers were genotyped as in previous works[9,14] or by PCR-RFLP methods developed in this work. The M12, M20, M65, M92, M107, M122, M148, M163, M165, M207, M269 and M377 (PA Underhill, pers. comm.) markers were amplified using previously published primers[10,15,16] and their allelic states diagnosed by means of the restriction enzymes *Nde*II, *Ssp*I, *Hin*fI, *Hpy*CH4IV, *Nla*III, *Mae*III, *Mnl*I, *Rsa*I, *Dra*I, *Mva*I and *Pst*I, respectively. Mismatched primers to create RFLPs were designed for M166 (Reverse 5′-CAGCG AATTAGATTTTCTTG-3′, digested with *Bsr*DI), M224 (Reverse 5′-TGAAATATTTGGAAGGGCTGAA-3′, digested with *Acu*I), M342 (Forward 5′-GTTAAATTATGACTTACGGG CA-3′, digested with *Bsp*1286I), P15 (Forward 5′-TGCTTGA GGTTCTGAATCATA-3′, digested with *Nde*I) and P16 (Reverse 5′-CCTGTCAATATTCCTGTTAAT-3′, digested with *Tsp*509I). Markers M124 and M175 were genotyped with published primers[10] by sequencing both strands (BigDye Terminator kit v.3.1) using an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Haplogroups were identified by following standardized nomenclature guidelines[13,17] (Figure 1). In order to increase the size of some population samples, in those populations for which more than one reference was found in the literature but which had been analyzed at different levels of genealogical resolution (SNP coverage), the frequency data described for a marker in the sample with less genealogical resolution, were subdivided according to the subgroup proportions observed in the more genealogically resolved sample of the same population. Both samples were then grouped together and hereafter considered as a single sample. Before applying this approach to a set of samples of the same population, congruency tests were carried out at the lowest common resolution level in order to check that the samples were not different at this common level. Only those samples that did not differ were subdivided at the next level of resolution determined by the more resolved sample.

STR markers

Our Basque sample had been previously typed for 11 Y-chromosome STRs[18] so that it has been possible to compare the STR variability within haplogroups both among our samples and among other North African, Iberian and European populations. In order to include the largest number of populations in the analyses only the following five STRs have been included in the present analysis: DYS19, DYS390, DYS391, DYS392 and DYS393 (Supplementary Information 2). We used this particular order for the STR-haplotype nomenclature. In addition, a subset of Iberian and Berber samples belonging to the relatively well represented European haplogroups R1b3*-M269, R1b3d-M153 and R1b3f-SRY2627 were genotyped for the above set of five STRs. Five-STR haplotypes from the same haplogroups, characterized in samples of diverse European and Near Eastern origin, have also been collected for comparative purposes. For R1b3d-M153, six chromosomes were included in the analysis, five of

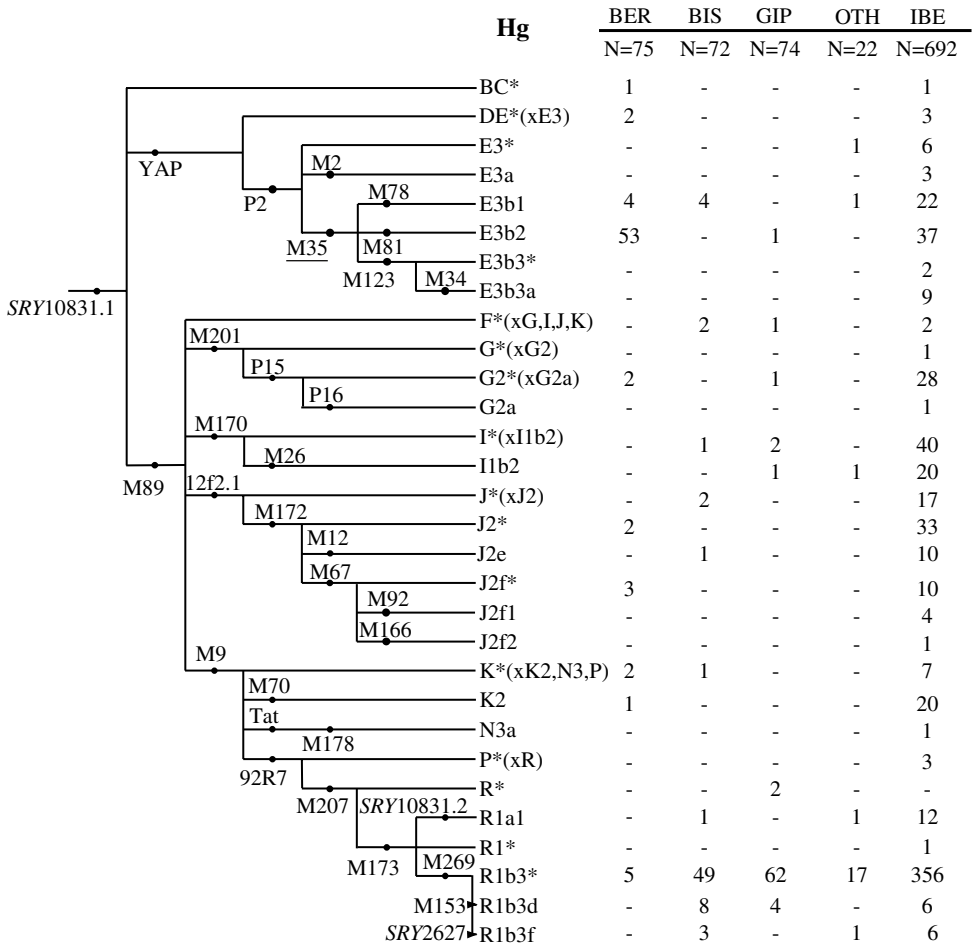

Figure 1 Genealogical relationships, nomenclature and frequencies of the Y chromosome haplogroups. Only informative markers are included. The status of the underlined marker was inferred. Markers M13, M20, M52, M65, M107, M122, M124, M148, M163, M165, M175, M224, M342 and M377 were also genotyped but not observed. BER: Berbers; BIS: Biscay Basques; GIP: Gipuzkoa Basques; OTH: Other Basques; IBE: non-Basque Iberians. Hg: Haplogroup.

which (Iberians) were genotyped in this work. For R1*(xR1a,R1b3f)-M173 a total of 3087 chromosomes were considered. In addition to the data from the literature, 27 R1*(xR1a,R1b3f)-M173 non-Basque Iberians and five R1*(xR1a,R1b3f)-M173 North Africans were genotyped in this work. Finally, a total of 57 chromosomes were analyzed for R1b3f-SRY2627. In addition to the data from the literature, one European (French) R1b3f-SRY-2627 sample was genotyped in this work.

Statistical analysis

Genetic diversities (Nei's heterozygosity, *h*, and the mean number of pairwise differences between five STR-loci haplotypes) were computed with ARLEQUIN 2.000.[19] Tests for significant pairwise differences in *h* were assessed by a Bayesian approach by means of TEST_h_DIFF (http://www.ucl.ac.uk/tcga/software/index.html) program under the MATLAB v.6.5 environment (The MathWorks Inc.). Reynolds F_{ST} genetic distances[20] between populations were also computed based on haplogroup or haplotype frequencies by means of ARLEQUIN 2.000. Multidimensional scaling (MDS) was used to represent genetic distances in two-dimensional space using SPSS ver. 11.5.1 (SPSS Inc.). To estimate the time to most recent common ancestor (TMRCA) of some subtrees we used Batwing[21] assuming constancy in population size or exponential growth ($\alpha = 0.005$/generation) and gamma priors for θ and ω. The mutation rates used were those described[22] for the specific loci used herein, except for those two loci with mean mutation rates of zero where, in order to be conservative, we decided to use the (higher) average rate described for Y-chromosome STR loci.[23] The locus-specific mean mutation rates[22] are higher (and therefore will provide younger age estimates) than the 95% upper limit of the generic rate[23] (mean + 2SD: 1.8×10^{-3}). Generation time was assumed to be 25 years.

Results

The Basque Y-chromosome gene pool

The majority (about 86%) of the Basque Y chromosomes belong to haplogroup R1*(xR1a,R1b3f)-M173, of which R1b3*-M269 accounts for 88% (Figure 1). As this haplogroup is also the most abundant type in all Western Europe[16] it places the Basque Y chromosomes within the European landscape. The above data are reflected in the low diversity values for the Basque populations (Table 1). Within R1b3*-M269, Basques also show a reduced STR diversity (Table 1). Thus, compared for instance to the non-Basque Iberians, the average number of mutations in our sampled Basques is significantly lower (Mann–Whitney *U*-test, $P=0.009$).

The presence of R1b3b-M65, a possibly autochthonous Basque branch of R, was not confirmed in our Basque samples. Instead, we did detect in our Basque sample the putative Iberian markers R1b3d-M153 and R1b3f-SRY2627 although in lower frequencies than in earlier analyses on Basques[11,24] (Figure 1). Thus, R1b3d-M153 shows a frequency in our total Basque sample of 7.1%, a figure higher than the corresponding frequency in non-Basque Iberians (0.9%). R1b3f-SRY2627 shows, in our Basque sample, a frequency of 2.4% (5.2% in Iberians). This haplogroup is considered to be of Iberian origin as the highest frequencies and diversities for R1b3f-SRY2627 have been described in the Mediterranean area of the Iberian Peninsula.[24,25]

The existence of a minor NW African male component (1.2%) in Basques is confirmed by the presence of the African lineages E3*-P2 and E3b2-M81. The latter is the most widespread haplogroup of the E cluster in Iberia, reaching its highest frequencies both in southern and northern parts of the Iberian Peninsula.[9,26]

Relationships among the Basque samples

To explore the hypothesis that the Basques might not constitute a homogeneous population,[27,28] we tested the pairwise F_{ST} values among the individual Basque samples. For the binary markers, after Bonferroni correction ($\alpha=0.008$), the hypothesis of a single Basque population is rejected (Supplementary Information 3). Sequential testing rendered the most inclusive three-sample group formed by Biscay, Gipuzkoa-2 and Other Basques (hereafter referred to as 'pooled Basques'). After Bonferroni correction of the pairwise comparisons between Gipuzkoa-1 with the samples forming this group ($\alpha=0.0167$), Gipuzkoa-1 remained significantly different. However, the possibility of structure among Basques is minimized after comparing all the Basque samples for their five-STR haplotype composition within the main haplogroup R1*(xR1a,R1b3f)-M173 (Supplementary Information 3). The modal haplotype is the same in the five samples (Biscay, Gipuzkoa-1, Gipuzkoa-2, Other Basques and the Basque sample from Brión *et al*[12]), being the number of

Table 1 Diversity values for the populations considered

		5-STR haplotypes within R1(xR1a,R1b3f)-M173*		
Haplogroups population	*h*±*SD*	*Population*	*h*±*SD*	*Mean NM*[a]
Biscay	0.5246±0.0685	Biscay	0.6995±0.0642	1.374
Gipuzkoa-1	0.2969±0.0692	Gipuzkoa-1	0.6115±0.0626	1.045
Gipuzkoa-2[b]	0.6374±0.0716	Gipuzkoa-2[b]	0.8138±0.0505	1.802
		Brión *et al* (2003)	0.8261±0.0747	1.692
Other Basques	0.4113±0.1308	Other Basques	0.6550±0.1115	0.889
Pooled Basques[c]	0.5468±0.0488	All Basques	0.7090±0.0332	1.345
Iberians	0.7171±0.0182	Iberians	0.8899±0.0109	2.034
French	0.6561±0.0406	Frisians	0.8348±0.0367	1.838
English	0.5154±0.0148	English	0.8728±0.0074	1.825
Scottish	0.4372±0.0217	Scottish	0.8809±0.0097	1.891
Irish	0.3048±0.0296	Irish	0.9174±0.0080	2.182
Welsh	0.3017±0.0419	Wales	0.8793±0.0144	1.786
Italians	0.7887±0.0195	Italians	0.9368±0.0427	2.589
Sardinians	0.8015±0.0139	Belgians	0.9183±0.0358	2.559
Sicilians	0.8310±0.0165	Icelandics	0.8757±0.0276	2.270
Dutch	0.6560±0.0316	Austrians	0.9164±0.0252	2.504
Germans	0.7064±0.0194	Germans	0.8853±0.0393	1.808
Czech-Slov.	0.7623±0.0204	Croatians	0.9715±0.0131	3.153
Poles	0.6488±0.0265			
Danes	0.6770±0.0171	Danes	0.8759±0.0227	1.781
Norwegians	0.6924±0.0073	Norwegians	0.8785±0.0213	1.894
Croatians	0.7251±0.0349			
Turks	0.8934±0.0050	Turks	0.9513±0.0110	3.259
Georgians	0.7243±0.0253	Armenians	0.9033±0.0141	2.681
Berbers	0.4951±0.0702	Berbers	0.9684±0.0205	4.498

[a]Mean number of mutations.
[b]Basques in Underhill *et al*.[10]
[c]All Basques except Gipuzkoa-1.

different haplotypes highest in Biscay (18/61). After Bonferroni correction ($\alpha = 0.005$) (Supplementary Information 3), all Basque samples showed nonsignificant differences for the five-STR haplotypes within the haplogroup. Therefore, the above differences between the samples from Gipuzkoa do not appear to be the result of different internal composition within R1*(xR1a,R1b3f)-M173, but to the higher proportion of this lineage within our Gipuzkoa-1 sample. Thus, while these data cannot be strictly taken as a proof of genetic structure within Basques,[29] it may however warn against taking any Basque sample as representative of the Basques.

Relationships between the Basque and other samples

It has been suggested that the British Celtic populations and the Basques are derived from common paternal ancestors and that genetic drift in these populations has not been sufficiently great to differentiate them.[6] In this regard, for haplogroups, the pooled Basques are more diverse than the samples from Ireland ($P < 0.0001$), Wales ($P < 0.0001$) and Scotland ($P = 0.04$), while the Gipuzkoa-1 sample does not show significant differences with these populations (*P*-values 0.88, 0.94, 0.054, respectively). In this context, pairwise comparisons (F_{ST} values) of the Basque samples with other European populations based on haplogroup frequencies show that Gipuzkoa-1 has its closest affinities with the Irish and Welsh (Supplementary Information 3). These similarities can be explained in terms of the R1*(xR1a,R1b3f)-M173 frequencies, which are highest in Gipuzkoa-1 (0.84) and Ireland (0.83). The pooled Basques (excluding Gipuzkoa-1) showed significant F_{ST} values with all the populations.

Within Western Europe, the low Basque haplogroup diversity stands out when compared to their geographical neighbors. Thus, for haplogroups, both the pooled Basque sample and the Gipuzkoa-1 sample are less diverse than the non-Basque Iberians ($P = 0.0004$ and < 0.0001, respectively). The overall landscape of haplotypic diversity within R1*(xR1a,R1b3f)-M173 (Figure 2) confirms that Basques are the least diverse of all populations. Basques, who share with Iberians and Italians the same Atlantic modal haplotype (14,24,11,13,13),[6] show an outlier position (Figure 3), in agreement with their low diversity values.

Among the Basques, the Gipuzkoa-1 sample is the least diverse (Table 1). However, in this case, F_{ST} comparisons between the global Basque population and the rest of the populations (using the five STR-loci variability within R1*(xR1a,R1b3f)-M173) (Supplementary Information 3) do not show special affinities between the Basques and the Irish or Welsh. Similarly, the diversity of Basques for the R1*(xR1a,R1b3f)-M173 associated STR haplotypes is

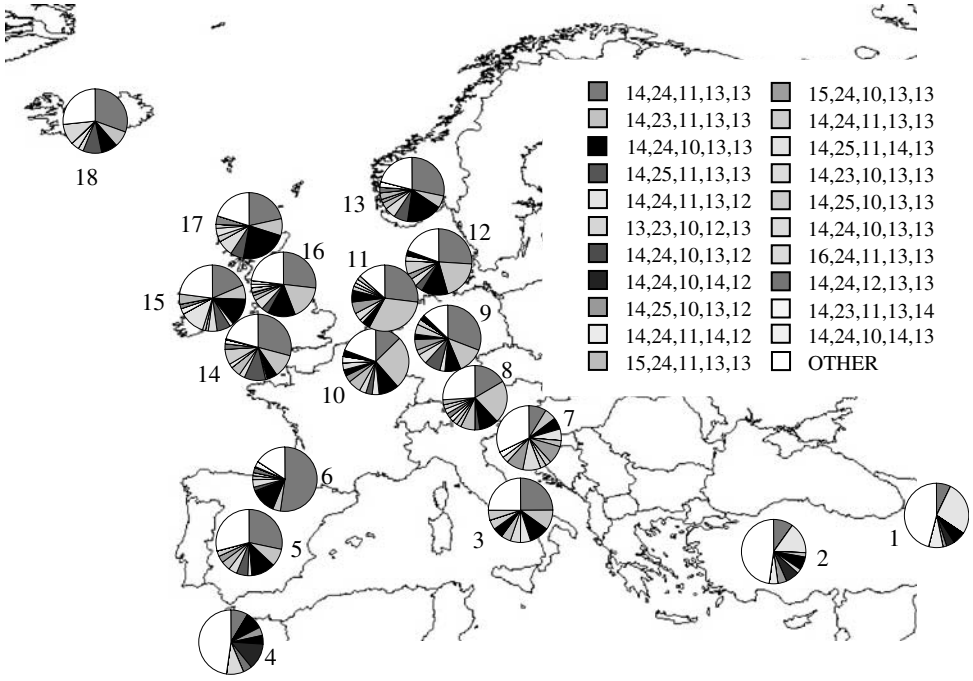

Figure 2 Map of the frequency distribution of the R1*(xR1a,R1b3f)-M173 STR loci haplotypes in Europeans, Near East and North Africa. For clarity, only those haplotypes with a frequency above 2% are indicated. Populations: 1: Armenians ($n = 238$); 2: Turkish ($n = 90$); 3: Italians ($n = 20$); 4: Berbers ($n = 23$); 5: Non-Basque Iberians ($n = 437$); 6: All Basques ($n = 209$); 7: Croatians ($n = 34$); 8: Austrians ($n = 42$); 9: Germans ($n = 37$); 10: Belgians ($n = 31$); 11: Friesians ($n = 52$); 12: Danes ($n = 77$); 13: Norwegians ($n = 113$); 14: Welsh ($n = 244$); 15: Irish ($n = 285$); 16: English ($n = 799$); 17: Scottish ($n = 370$); 18: Icelanders ($n = 75$). Haplotype nomenclature refers to alleles of loci DYS19, DYS390, DYS391, DYS392 and DYS393 in that order.

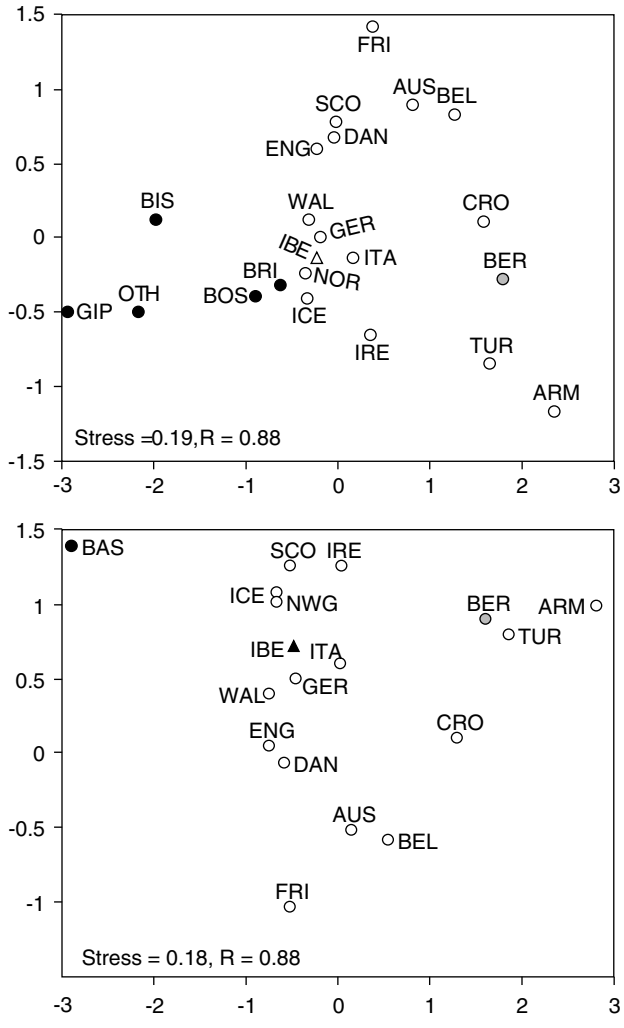

Figure 3 Multidimensional Scaling Analysis of the R1*(xR1a,R1b3f)-M173 STR loci haplotypes in Europeans, Near East and North Africa. ARM: Armenians; AUS: Austrians; BAS: All Basques grouped; BEL: Belgians; BER: Berbers; DAN: Danes; ENG: English; FRI: Friesians; GER: Germans; IBE: Iberians; IRE: Irish; ITA: Italians; NWG: Norwegians; SCO: Scottish; TUR: Turkish; WAL: Welsh; CRO: Croatians; ICE: Icelanders. Samples typed in this work are indicated as follows: black circles: Basques; gray circles: Berbers; black triangle: Iberians.

significantly different from that of Iberians, Irish, Welsh and Scottish (all $P<0.001$), whose diversity values are among the highest.

This contrasting pattern between haplogroup diversity and the R1*(xR1a,R1b3f)-M173 associated STR diversity between Basques, on the one hand and Irish and Welsh on the other, can be graphically observed in Figure 4. Thus, while Basques show a proportional reduction in STR diversity for their low haplogroup diversity values, the British population in general and Wales and Irish in particular, show, even for their low binary diversity, a 'saturation' (steady-state) level of STR haplotype diversity.

The age of the Basques

An approximation to infer the age of a specific population is based on the estimating the age of haplogroups that originated within that geographical area. As we have not detected R1b3b-M65 in Basques, the best candidates left are R1b3d-M153 and R1b3f-SRY2627. Given that R1b3f-SRY2627 has higher STR haplotype h in Iberians (0.83 ± 0.06) than in Basques (0.73 ± 0.08), although this difference is not significant ($P=0.11$), and the average number of mutations is also higher in Iberians (1.7) than in Basques (1.3), the most plausible hypothesis is that this haplogroup originated in Iberians (Figure 5). R1b3d-M153 STR-haplotype diversity is not significantly different between Basques ($h=0.66\pm0.13$) and non-Basque Iberians ($h=0.60\pm0.23$) ($P=0.6$), and also, the number of different lineages in M153 Basques (seven out of 17 total) is similar to that in Iberians (three out of six total). However, the average pairwise mutational difference is two-fold in Basques (1.4 *vs* 0.7 in Iberians) (Figure 5). This could be indicative of a Basque origin for this haplogroup, particularly given that the sample size of Iberians is four-times larger (Figure 1) and more geographically widespread. The fact that this haplogroup is absent in the sample 'Other Basques' does not contradict this point, as from the binomial distribution, even if the real frequency of this haplogroup in that population was as high as 12% we could still score zero observations of this haplogroup in a sample of 22 individuals with $P=0.05$. On the other hand, even in the less likely scenario that the mutation originated somewhere else and was introduced into the Basque Country by migration, from the statistical point of view the introduction of an allele by migration and the introduction of an allele by mutation are equivalent concepts. However, given the small frequency of this haplogroup outside the Basque Country we favor a Basque origin for this haplogroup. Alternatively, R1b3d-M153 may be present in a common ancestral population, arising to relatively higher frequency in Basques through genetic drift. However, this scenario would have led also to a reduction in STR diversity of R1b3d-M153 in Basques, which is not supported by our data.

Thus, under this assumption, the TMRCA of R1b3d-M153 could be taken as a lower bound for the age of the Basque population. Batwing simulations, using parameters obtained as described in Figure 6, indicate that the ages range between 17 900 (10 700–26 500) years, assuming exponential growth, and 21 300 (8500–51 000) years, under constant population size. Therefore, these ages indicate that this population, or at least some of its Y chromosome lineages, dates back to pre-Neolithic times. This estimate is supported by inferences made using highly variable autosomal minisatellite loci.[30]

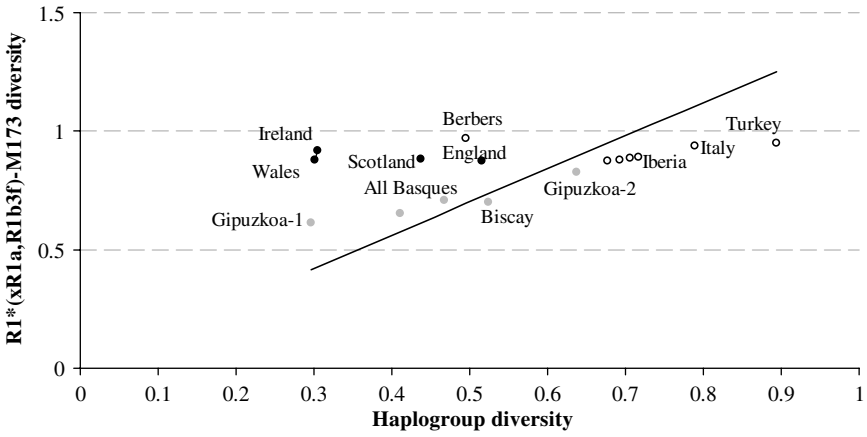

Figure 4 Haplogroup *h* (*x*-axis) *vs* R1*(xR1a,R1b3f)-M173 STR loci haplotype diversity (*y*-axis). We choose R1*(xR1a,R1b3f)-M173 because it is the major haplogroup in Basques and the British populations. For clarity, not all population names are indicated. Black circles: British populations; gray circles: Basque populations; empty circles, rest of the populations. Graph includes a linear regression line.

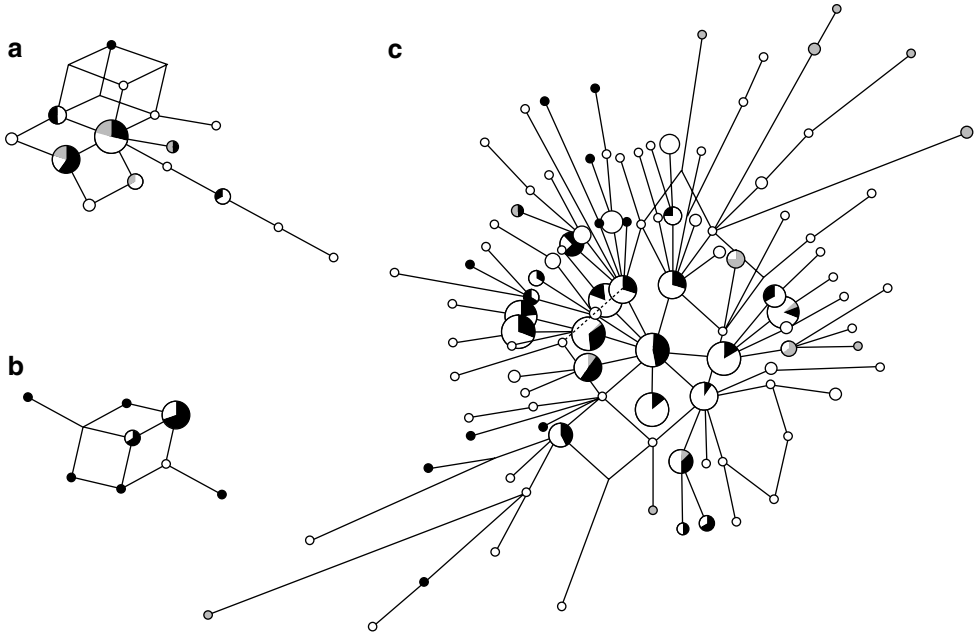

Figure 5 Networks representing the diversity patterns for the Basque and other populations' STR haplotypes within defined haplogroups. (**a**) STR haplotypes within R1b3f-SRY2627, (**b**) within R1b3d-M153 and (**c**) within R1b3*-M269. Black circles represent Basque haplotypes and white circles, non-Basque Iberian haplotypes. In (**a**) and (**b**), gray circles represent European haplotypes but in (**c**) they represent Berber haplotypes. Circles areas are proportional to frequency. Haplotype relationships were obtained applying sequentially reduced median and median joining methods implemented in Network 4.0.[36]

Discussion

Combined haplotype information of slow (SNPs) and more quickly (STRs) evolving markers can be used to reveal a greater detail about the demographic and evolutionary processes that have played a role in the history of human populations. However, a note of caution must be added to reflect the fact that, as it happens with mtDNA, herein we are focusing on just a single locus, which may not have been immune to the effects of selection. The analysis at the haplogroup level in the set of (mainly) European populations shows a marked drop in diversity in the Basque populations. This happens despite Basques having been represented in the Y-chromosome SNP-discovery panel of samples, a fact that through ascertainment bias may produce a higher diversity than real for the populations in the discovery panel. The low observed diversity causes a certain affinity between the Basques and the populations of the British Isles, particularly Irish and Welsh. However, this

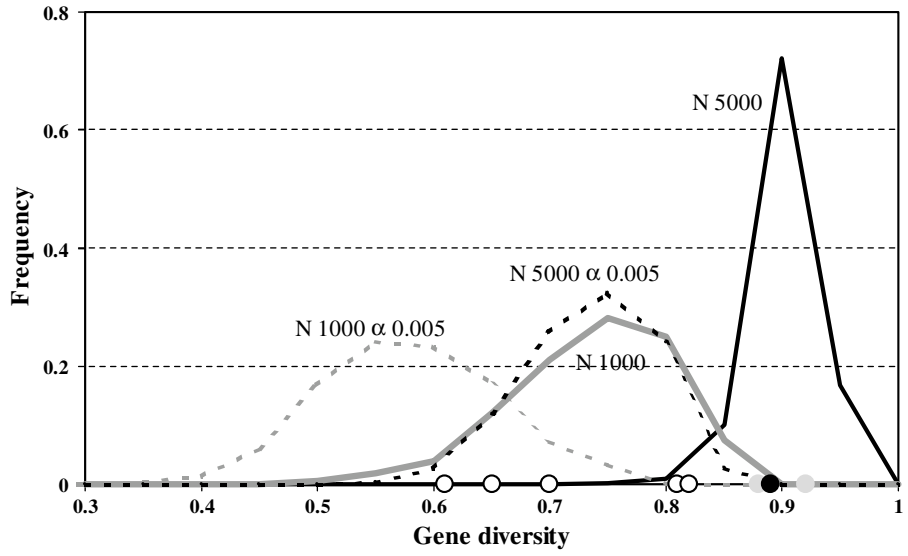

Figure 6 Expected diversity distributions using coalescent simulations. For the determination of proper estimates for the effective populations size and population growth coefficient for input in Batwing, coalescent simulations were run by means of Simcoal2. The evolution of five-STR haplotypes was simulated using the mutation rates described[22] (see main text). For those loci with an average mutation rate of zero the average Y-chromosome STR rate of 0.0007 per generation was used.[23] Simulations assumed a stepwise mutation model with geometric parameter of 0.5 and range constraint of 25. Simulations indicate that a constant effective population size of 1000 or a population of effective size of 5000 growing with an exponential rate of 0.005 per generation can satisfactorily explain the diversity levels observed in the Basque samples These parameter estimates were later used to estimate the age of the R1b3d-M153 lineage with Batwing. In this case, the effective population size was reduced proportionally to the frequency of this haplogroup (approx. 1/10). Continuous lines represent simulations with constant population size. Dashed lines, simulations with an exponential growth rate of 0.005/generation. Black lines $N_e = 5000$; gray lines $N_e = 1000$. White circles on the *X*-axis: Basque samples (from left to right: Gipuzkoa-1, Other Basques, Biscay, Gipuzkoa-2, Brión *et al*[12] Basques); gray circles, British sample (the circle on the left represents any of the English, Scottish or Welsh samples; the circle on the right represents the Irish sample); black circle, Iberian sample.

may be simply the effect of convergent drift, as when we consider the STR haplotypes within the major haplogroup (R1b) the latter populations are no more closely connected to the Basques than other European populations. Particularly Irish and Welsh show much higher diversities within R1b than Basques (and Gipuzkoa-1 in particular). The high associated STR diversity points to a prehistoric founder effect for the Welsh and the Irish, long enough ago that the more quickly evolving STR diversity has regenerated. Our own simulations with Simcoal2 demonstrate that a splitting founding population of $N_e = 50$ growing at an exponential rate of $\alpha = 0.005$ can regenerate its five-STR haplotype diversity in 400 generations (assuming a source population with N_e of 5000) (not shown).

Basques share with the rest of Europeans both the most common haplogroup (R1*(xR1a,R1b3f)-M173) and the modal STR haplotypes within this haplogroup.[15] The low STR diversity in Basques seems to be the result of a lower effective population size maintained through generations, which is particularly marked in Gipuzkoa-1. This low effective size may have allowed drift to drive some haplogroups to such high frequencies. It can also be argued that at least part of this conspicuously low diversity present in Basques can be attributed to a sampling bias. One of the criteria for donors to be included in the sample is to have at least four generations of Basque ancestry (recorded by Basque surnames) plus, in many cases, a localized ancestry of their grandfathers within the district being sampled. Stringent criteria can lead to reduced diversity, as was the case with the Gaelic surnames.[31] As this stringent criterion is not normally demanded in other sampling schemes, we can be introducing a reduction in the effective size of the population being sampled. First, we are discarding any contribution by external gene flow that may have take place during approximately the last 100 years (a restriction that is not normally imposed on other samples) and second, we are removing any internal gene flow (among Basque districts) that may have taken place during that time. In fact, other Basque samples[10–12] do not show such an important drop in diversity as seen in our Basque samples, although still show slightly lower values both in haplogroups and within R1*(xR1a,R1b3f)-M173 than, for instance, Iberians (Figure 5).

In any case, this low effective size is not the result of a recent founder effect, as our data support the hypothesis that at least some lineages of Y chromosome in modern Basques originated and have been evolving since pre-Neolithic times. We cannot gauge up to which point the origin and evolution of these lineages has been geographically local, but this possibility should be unsurprising given that there is evidence supporting human presence in the Basque Country since the Lower Paleolithic, about 150 000

years ago, although the oldest skeletal remains found correspond to the Neanderthals (in the Middle Paleolithic). As regards archaeological sites in the Basque Country,[32] the Upper Paleolithic is one of the richest periods with some of the sites showing continuity in habitation up to, at least, the Bronze Age (about 2000 BC). However, it can be argued that Archaeology can seldom differentiate between the cultural and/or biological evolution of a single group and the possible replacement by new groups of incomers. Ancient DNA analysis focused on the Y chromosome could yield the proof needed to conclude a local evolution of Basques.

There is some evidence of a short-range outward flow of Basque Y chromosomes, as the presence of R1b3d-M153 chromosomes in Iberia suggests. This finding is in agreement with previous data,[24] which provided more evidence for such gene flow between Basques and surrounding populations on the basis of haplogroup R1b3f-SRY2627. However, in agreement with additional data,[33] our data do not show any signs of long range diffusion of Basque Y-chromosome haplogroups into North Europe associated to the retreat of the last Glacial Maximum, as has been suggested for mitochondrial DNA.[34,35] Finally, while a pre-Neolithic settlement for the Basques can be posited, the strong genetic drift experienced by the Basques does not allow to consider Basques either the only or the best representatives of the ancestral European gene pool. Similarly, genetic drift will make determination of their population affinities difficult.

Acknowledgements

CF is a FUNCIS Postdoctoral Fellow. SA is a Ramón y Cajal Fellow. This work has been partly funded by Ministerio de Educación y Ciencia and by the UPV/EHU project 9/UPV 00154.310-14495/2002 (to CR) and by grants BMC2001-3511 from Ministerio de Ciencia y Tecnología and COF2002-015 from Gobierno de Canarias to VMC.

References

1 Barbujani G: Genes, people and languages. *Am J Hum Genet* 2000; **67**: 264–266.
2 Etcheverry MA: El factor rhesus: su genética e importancia clínica. *El Día Médico* 1945; **17**: 1237–1259.
3 Mourant AE: The blood groups of the Basques. *Nature* 1947; **160**: 505.
4 Bertranpetit J, Cavalli-Sforza LL: A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann Hum Genet* 1991; **5**: 51–67.
5 Gorrochategui J: Planteamiento de la linguística histórica en la datación del euskara; in XV Congreso de Estudios Vascos Donostia/San Sebastian: Eusko Ikaskuntza/Sociedad de Estudios Vascos, 2002, pp 103–114.
6 Wilson JF, Weiss DA, Richards M, Thomas M, Bradman N, Goldstein DB: Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* 2001; **98**: 5078–5083.
7 Chikhi L, Nichols RA, Barbujani G, Beaumont MA: Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci USA* 2002; **99**: 11008–11013.
8 González AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM: Mitochondrial DNA Affinities at the Atlantic Fringe of Europe. *Am J Phys Anthropol* 2003; **120**: 391–404.
9 Flores C, Maca-Meyer N, González AM *et al*: Reduced genetic structure of the Iberian Peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet* 2003; **12**: 855–863.
10 Underhill PA, Shen P, Lin AA *et al*: Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000; **26**: 358–361.
11 Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J: 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* 2001; **68**: 1019–1029.
12 Brión M, Salas A, González-Neira A, Lareu MV, Carracedo A: Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite. *Am J Phys Anthropol* 2003; **122**: 147–161.
13 Jobling MA, Tyler-Smith C: The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 2003; **4**: 598–612.
14 Flores C, Maca-Meyer N, Pérez JA, González AM, Larruga JM, Cabrera V: A predominant European ancestry of paternal lineages from Canary islanders. *Ann Hum Genet* 2003; **67**: 138–152.
15 Semino O, Passarino G, Oefner PJ *et al*: The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 2000; **290**: 1155–1159.
16 Cinnioðlu C, King R, Kivisild T *et al*: Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 2004; **114**: 127–148.
17 Y Chromosome Consortium: a nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; **12**: 339–348.
18 García O, Martín P, Gusmao L *et al*: A Basque Country autochthonous population study of 11 Y-chromosome STR loci. *Forensic Sci Int* 2004; **145**: 65–68.
19 Schneider S, Roessli D, Excoffier L: *Arlequin Ver. 2.000: a software for population genetics data analysis*. Switzerland: Genetics and Biometry Laboratory, University of Geneva, 2000.
20 Reynolds JB, Weir S, Cockerham CC: Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 1983; **105**: 767–779.
21 Wilson IJ, Weale ME, Balding DJ: Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J Roy Statist Soc A* 2003; **166**: 155–201.
22 Kayser M, Roewer L, Hedman M *et al*: Characteristics and frequency of germline mutations at *STR* loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 2000; **66**: 1580–1588.
23 Zhivotovsky LA, Underhill PA, Cinnioglu C *et al*: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004; **74**: 50–61.
24 Hurles ME, Veitia R, Arroyo E *et al*: Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-Chromosomal DNA polymorphism. *Am J Hum Genet* 1999; **65**: 1437–1448.
25 Rosser HZ, Zerjal T, Hurles ME *et al*: Y-Chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 2000; **67**: 1526–1543.
26 Maca-Meyer N, Sánchez-Velasco P, Flores C *et al*: Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet* 2003; **67**: 329–339.
27 Aguirre AI, Vicario A, Mazón LI *et al*: Are the Basques a single and a unique population? *Am J Hum Genet* 1991; **49**: 450–458.
28 Iriondo M, Barbero MC, Manzano C: DNA polymorphisms detect ancient barriers to gene flow in Basques. *Am J Phys Anthropol* 2003; **122**: 73–84.

29 Cavalli-Sforza L, Feldman MW: Spatial subdivision of populations and estimates of genetic variation. *Theor Popul Biol* 1990; **37**: 3–25.
30 Alonso S, Armour JAL: MS205 minisatellite diversity in Basques: evidence for a pre-Neolithic component. *Genome Res* 1998; **8**: 1289–1298.
31 Hill EW, Jobling M, Bradley DG: Y-chromosome variation and Irish origins. *Nature* 2000; **404**: 351–352.
32 De la Rúa C: Los estudios de paleoantropología en el País Vasco. *Munibe* 1990; **42**: 199–219.
33 Izagirre N, de la Rúa C: 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* 1999; **65**: 199–207.
34 Torroni A, Bandelt HJ, D'Urbano L *et al*: mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 1998; **62**: 1137–1152.
35 Achilli A, Rengo C, Magri C *et al*: The molecular dissection of mitochondrial DNA haplogroup H confirms that the franco-cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 2004; **75**: 910–918.
36 Bandelt HJ, Forster P, Röhl A: Median-joining networks for inferring intraespecific phylogenies. *Mol Biol Evol* 1999; **16**: 37–48.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)