

Y Chromosome and Mitochondrial DNA Characterization of Pasiegos, a Human Isolate from Cantabria (Spain)

N. Maca-Meyer¹, P. Sánchez-Velasco², C. Flores³, J.-M. Larruga¹, A.-M. González¹, A. Oterino⁴ and F. Leyva-Cobián^{2*}

¹Departamento de Genética, Facultad de Biología, Universidad de La Laguna, Tenerife, Spain

²Servicio de Inmunología, Hospital Universitario "Marqués de Valdecilla", Servicio Cántabro de Salud, Santander, Spain

³Unidad de Investigación, Complejo Hospitalario "Nuestra Señora de Candelaria", Servicio Canario de Salud, Tenerife, Spain

⁴Servicio de Neurología, Hospital Universitario "Marqués de Valdecilla", Servicio Cántabro de Salud, Santander, Spain

Summary

Mitochondrial DNA sequences and Y chromosome haplotypes were characterized in Pasiegos, a human isolate from Cantabria, and compared with those of other Cantabrian and neighbouring Northern Spain populations. Cantabria appears to be a genetically heterogeneous community. Whereas Lebaniegos do not differ from their eastern Basque and western Asturian and Galician neighbours, Pasiegos and other non-Lebaniego Cantabrians show significant differences with all of them. Pasiegos are peculiar for their high frequencies of Y chromosomal markers (E-M81) with North African assignation, and Y chromosomal (R-SRY2627) and mtDNA (V, I, U5) markers related to northern European populations. This dual geographic contribution is more in agreement with the complex demographic history of this isolate, as opposed to recent drift effects. The high incidence in Cantabrians with pre-V and V mtDNA haplotypes, considered as a signal of Postglacial recolonization in Europe from south-western refugees, points to such refugees as a better candidate population than Basques for this expansion. However, this does not discount a conjoint recolonization.

Keywords: Y-Chromosome; mtDNA; Pasiegos

Introduction

Although no separate ethnic group is considered to exist in the north of Spain, the idea of an archaic lifestyle in the North was coined in the past and to a certain degree is still maintained. From a geographical viewpoint, what the people living in these areas have in common is a landscape of valleys, separated from each other by forests and mountains, which partly explains some of the characteristics classically attributed to them. Cantabria is one of these northern mountainous regions (geographi-

cally, the north of Spain is also known as the Cantabrian Cornice). It has great anthropological interest because there are vestiges of continuous human occupation of this area since Palaeolithic times (Fernández-Acebo, 1996). Probably the best worldwide renowned artistic manifestations of that period are the Aurignacian paintings of the Altamira and Puente Viesgo caves (Bischoff *et al.* 1992). Due to the abrupt orography of the region and the ferocity of the tribal inhabitants, Cantabria was rather impermeable to the numerous waves of invaders that successively colonised the Iberian peninsula until recent historic times (González-Echegaray, 1993). Although the present community have not conserved an ancient language, as have their eastern neighbours the Basques, there are interesting human isolates in Cantabria, the Pasiego being the most characteristic one

*Correspondence to: Francisco Leyva-Cobián, Servicio de Inmunología, Hospital Universitario "Marqués de Valdecilla", Servicio Cántabro de Salud, 39008-Santander, Spain Telephone: 34-942-202549; Fax: 34-942-202726. E-mail: fleyvaco@humv.es

(Tax-Freeman, 1979). They are shepherds from Montes de Pas that seasonally move along the mountains carrying their livestock to the best pastures. The region's actual isolation is confirmed by a high degree of endogamy and consanguinity (Tax-Freeman, 1979; Pérez-Flores, 1998; Sánchez-Velasco *et al.* 1999), and a relatively high incidence of some diseases, such as multiple sclerosis, and virtual absence of others such as insulin-dependent diabetes mellitus. Furthermore, genetic studies using GM and KM immunoglobulin markers demonstrated a high degree of genetic differentiation of Pasiegos, that when included with other populations, reveals a considerable range of genetic variation within the Iberian peninsula, higher than that often polarised around the Basque versus non-Basque variation (Esteban *et al.* 1998). Concerning the origin of the Pasiegos at least three theories can be considered: i) Pasiegos are descendants of the primitive inhabitants of Cantabria. This hypothesis tacitly implies that human settlements in Montes de Pas have been quite constant since prehistoric times with no obvious population hiatus (Fernández-Acebo, 1996), and that the historically documented award of pastures to the Oña monastery in the eleventh century (Fernández-Acebo, 1996; González-Echegaray, 1993) did not substantially affect the demographic evolution of the indigenous settlers. ii) They are Goths that retreated into these valleys after the Moslem invasion. Congruent with this second hypothesis is a study on MHC class II polymorphisms that closely relate Pasiegos to Danes, Poles and Germans rather than to non-Pasiego Cantabrians (Sánchez-Velasco *et al.* 1999, 2003). iii) Pasiegos are descendants of Jews and Moors that took refuge in the area after their expulsion from Christian communities. In favour of this hypothesis is the high frequency found in

a small Pasiego sample of a Y-chromosome haplogroup with a likely North African Berber origin (Scozzari *et al.* 2001).

In this study we use maternally inherited mitochondrial DNA (mtDNA) and paternally inherited Y-chromosome polymorphisms to provide new information on the genetic structure of the Pasiego isolate, and to relate them to surrounding Iberian populations as well as other European and North African populations that share haplotypes with a well established geographic origin.

Materials and Methods

Samples

A total of 242 unrelated individuals from the Cantabrian region were analysed in this study. All potential donors gave informed consent, but for those from isolated areas only those with female and male ancestors born in the same place for at least two generations were taken into account. The geographic origins of the samples (Fig. 1) were as follows: 82 Pasiegos from any of the Three Pasiego Villas at the southern border of the current province of Cantabria, 72 Lebaniegos (inhabitants from Potes and surround, a small western Cantabrian locality near to Asturias), and a pool of 88 non-Pasiego and non-Lebaniego Cantabrians from different localities. In addition, published data from nearby populations were incorporated in comparative analyses: 45 Spanish Basques (Underhill *et al.* 2000) 90 Asturians and 19 additional Pasiegos (Scozzari *et al.* 2001) were used for Y-chromosome comparisons, and 135 Galicians and 167 Basques, as detailed elsewhere (Larruga *et al.* 2001), were used in the mtDNA analysis.

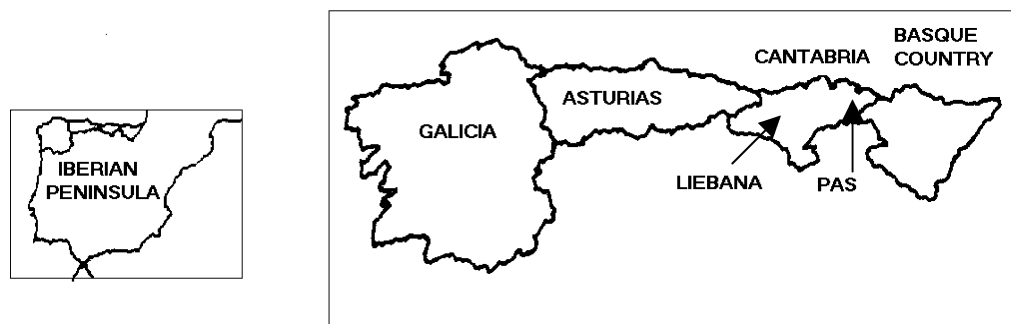


Figure 1 Sketch maps of the Iberian Peninsula (left) and the present provincial boundaries in the Cantabrian Cornice (right) with approximate geographic location of the populations sampled.

DNA was extracted from 5 μ l of blood using an alkaline protocol (Rudbeck & Dissing, 1998).

Binary Y-Chromosome Polymorphisms

Y chromosomes were typed using 25 biallelic polymorphisms: A-M13C, DE-YAP(Alu+), E-M2G, E-M78T, E-M81T, E-M123A, E-M34T, E-P2T, F-M89T, G-M201T, H-M52C, I-M170C, I-M26A, J-12f2 (8kb), J-M172G, J-M12T, K-M9G, K-M70C, N-TatC, P-92R7T, R-M173C, R-M153A, R-SRY2627T and R-SRY10831bA. Following the mutation-based nomenclature of the YCC (Y Chromosome Consortium, 2002), the first capital letter/s of each marker defines the haplogroup in which this marker is included. The last capital letter defines the derived state of the precedent marker. In the case of YAP, Alu+ indicates the presence of an Alu element, and, in the case of 12f2, 8kb indicate the absence of a LINE element, tested by the presence of a PCR fragment of 2078 bp (Sun *et al.* 2000). Primers used to amplify these markers and references for their methods of detection are available (Y Chromosome Consortium, 2002). The phylogenetic relationship among these markers is represented in Figure 2.

MtDNA Sequencing and Restriction Analyses

A sequence of 403 bp of the first hypervariable segment of the mtDNA control region (HV1) from position 15997 to 16399 (Anderson *et al.* 1981) was determined and sorted into defined haplogroups (Richards *et al.* 2000). To confirm this HV1-based haplogroup classification, all individuals assigned to a specific haplogroup were additionally tested by restriction analysis of the diagnostic coding-region mutations proposed to unambiguously classify sequences into haplogroups (Richards *et al.* 2000). Those few sequences whose HV1 and restriction motif did not match were subsequently tested for alternative haplogroups until fully classified.

Data Analyses

Statistical analyses were carried out using the Arlequin 2000 software package (Schneider *et al.* 2000). The level of genetic structure and relative affinities be-

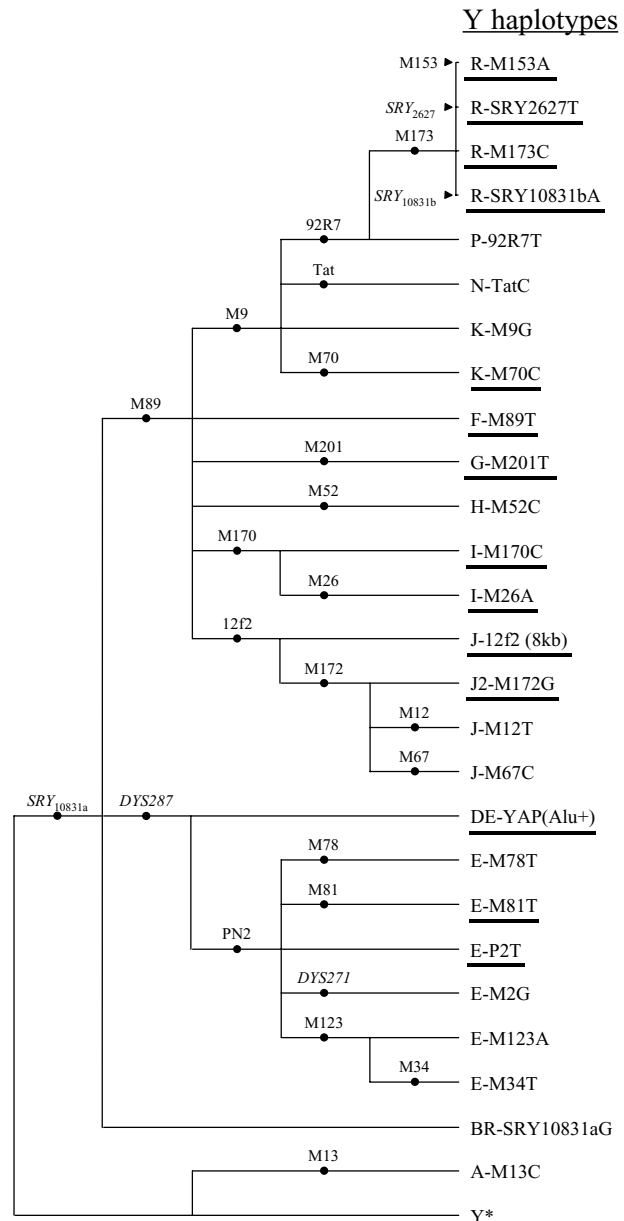


Figure 2 Most parsimonious tree of surveyed Y haplotypes. Markers are shown on branches. Haplotypes found are underlined.

tween the populations analysed were assessed using F_{ST} statistics to compare haplogroup frequencies and haplotypic frequencies for mtDNA and Y-chromosome markers respectively. The permutation procedure (10,000 permutations) was used to test the level of significance for the F_{ST} values. In addition, we used the formula:

$$i = \frac{H_{XY}}{\sqrt{H_X H_Y}}$$

(where H_{XY} is the number of haplotypes shared between populations X and Y, H_X is the number of different haplotypes present in population X, and H_Y is the number of different haplotypes present in population Y) as an estimation of identity (i) between populations. In this measure, we have omitted molecular differences under the supposition that new mutations have not occurred during the time after the populations split. We also suppose that selective forces are negligible, and that the relative similarity between populations is only due to the counteracting forces of gene flow and genetic drift due to founder effects and different demographic histories. Under these conditions, the number of different haplotypes in a population is a more reliable parameter than the relative frequency of haplotypes. For the same reasons, we have chosen the proportion of the number of different haplotypes to the total number of individuals analysed in a population as a measure of haplotype diversity.

Results

Y-Chromosome

Haplotype frequencies present in the Cantabrian populations studied are reported in Table 1. As is typical in all European populations, around 50% of haplotypes in the three samples belong to R-M173C. The absence in Lebaniegos of the North African E-M81T marker and the R-SRY10831bA derivative of the European R*M173C marker contrasts with their high frequencies in Pasiegos (18% in both markers) and in the pooled

Table 1 Y-chromosome haplotype frequencies in Cantabrian populations

Haplotype	Pasiegos	Lebaniegos	Other Cantabrians
DE-YAP(Alu+)	–	2 (0.05)	–
E-M81T	8 (0.18)	–	6 (0.17)
E-P2T	–	–	1 (0.03)
F-M89T	1 (0.02)	–	–
G-M201T	3 (0.07)	4 (0.11)	2 (0.05)
I-M170C	1 (0.02)	2 (0.05)	1 (0.03)
I-M26A	–	–	1 (0.03)
J-12f2 (8kb)	–	3 (0.08)	–
J-M172G	2 (0.04)	1 (0.03)	1 (0.03)
K-M70C	–	3 (0.08)	–
R-M173C	21 (0.47)	17 (0.46)	18 (0.50)
R-M153A	–	–	1 (0.03)
R-SRY2627T	1 (0.02)	5 (0.14)	2 (0.05)
R-SRY10831bA	8 (0.18)	–	3 (0.08)

sample of Cantabrians (17% and 8% respectively). On the other hand, the abundance of R-SRY2627T, another European R*M173C derivative, in Lebaniegos compared to Pasiegos (2%) and pooled Cantabrians (5%), and the absence of K-M70C in these last two samples and its presence in the former (8%) also deserves mention. Common Middle Eastern haplotypes such as J-12f2(8kb) and G-M201T are also more frequent in Lebaniegos. Finally, low frequencies of DE-YAP(Alu+) African derivatives have been detected in Lebaniegos and in the pooled sample of Cantabrians.

Slatkin linearized F_{ST} s show that differences in haplotype frequencies reach significance ($P = 0.04$) only between Pasiegos and Lebaniegos, whereas the haplotypic diversity (measured as the proportion of the number of different haplotypes to the total number of individuals analysed) has decreasing values ranging from pooled Cantabrians (0.28) and Lebaniegos (0.22) to Pasiegos (0.18). After some haplotype pooling (Table 2) it is possible to compare Cantabrian samples to their eastern (Basque) and western (Asturian) neighbours. The addition of data from a small sample of Pasiegos, previously published (Scozzari *et al.* 2001), increases the frequency of the North African marker E-M81T to 24% in the total sample. Basques share with Lebaniegos a comparatively high frequency for R-SRY2627T (11%), and Basques (78%) and Asturians (59%) show the largest frequencies in the region for the most common European marker (R*M173C*). Also notable in Basques is the high frequency (16%) of R-M153A contained within this lineage (Underhill *et al.* 2000). In turn, Asturians stand out for the high incidence (11%) of YAP+ derivatives other than E-M81T or E-M2G. Furthermore, Asturians share with Lebaniegos the presence of J-12f2 (8kb) representatives.

The matrix of Slatkin linearised F_{ST} s (Table 3) shows that Pasiegos are a discrepant population, being significantly different from all samples excepting pooled Cantabrians. Besides Pasiegos, Basques and Asturians/Lebaniegos are also heterogeneous populations.

MtDNA

The mtDNA sequences found in the Cantabrian samples are presented in Table 4 sorted into haplogroups. H and, within it, the Cambridge Reference Sequence (CRS)(Anderson *et al.* 1981), are the most common

Table 2 Y-chromosome haplotype frequencies in Northern Spain

Haplotype	AST ¹ (n = 90)	PAS ² (n = 64)	LEB ³ (n = 37)	CAN ⁴ (n = 36)	BAS ⁵ (n = 45)
DE-YAP(Alu+)* ⁶	10 (0.11)	–	2 (0.05)	1 (0.03)	–
E-M81T	2 (0.02)	15 (0.24)	–	6 (0.17)	1 (0.02)
F-M89T* ⁷	10 (0.11)	7 (0.12)	6 (0.16)	4 (0.11)	4 (0.09)
J-12f2(8kb)	3 (0.03)	–	3 (0.08)	–	–
J-M172G* ⁸	6 (0.07)	2 (0.03)	1 (0.03)	1 (0.03)	–
K-M9G* ⁹	1 (0.01)	–	3 (0.08)	–	–
R-M173C* ¹⁰	53 (0.59)	27 (0.42)	17 (0.46)	19 (0.53)	35 (0.78)
R-SRY2627T	1 (0.01)	1 (0.01)	5 (0.14)	2 (0.05)	5 (0.11)
R-SRY10381bA	4 (0.04)	12 (0.18)	–	3 (0.08)	–

¹Asturians; ²Pasiegos; ³Lebaniegos; ⁴Other Cantabrians; ⁵Basques.

⁶Includes all individuals typed as YAP+ other than E-M81T and E-M2G.

⁷Includes individuals typed as F-M89T, G-M201T, I-M170C and I-M26A of this work and 2.1 and 2.2 of Scozzari *et al.* (2001).

⁸Includes individuals typed as J-M172G and J-M67C of this work and 9.2 and 9.3 of Scozzari *et al.* (2001).

⁹Includes individuals typed as K-M9G and K-M70C.

¹⁰Includes individuals typed as R-M173C and R-M153A.

Table 3 Pair-wise linearized F_{ST} between populations based on Y-chromosome haplotype frequencies

	AST	PAS	LEB	CAN	BAS
AST	–	0.067***	0.018	0.010	0.044*
PAS		–	0.063**	0.000	0.166***
LEB			–	0.016	0.089**
CAN				–	0.067*

*P < 0.05; **P < 0.01; ***P < 0.001

Population codes as in Table 2.

haplogroup and the most common haplotype respectively in the three samples. However, the highest frequencies for both types are among Lebaniegos. On the contrary, highest frequencies of haplogroup V haplotypes are found in Pasiegos (16%) and pooled Cantabrians (15%) compared to Lebaniegos (11%). There are also noticeable differences for some haplogroups with low global frequencies in Europe. For example, Pasiegos lack J sequences, a haplogroup considered to be a Neolithic signature in Europe (Richards *et al.* 1996), but show comparatively higher frequencies (6%) for haplogroup I, which has its highest frequencies in North and West Europe (Helgason *et al.* 2000), compared to pooled Cantabrians (1%) or Lebaniegos (3%). On the other hand, haplogroup U representatives are far more frequent in Pasiegos (21%) and pooled Cantabrians (24%) than in Lebaniegos (8%). Finally, it is worth mentioning the presence of African sequences in the Pasiego and pooled Cantabrians. A North African U6

sequence that is very common in Argelian Berbers (Corte-Real *et al.* 1996), and also detected in Northern Portugal (González *et al.* 2003), has been found in pooled Cantabrians. Besides, pooled Cantabrians harbour a peculiar M1 sequence characterised by a back mutation in position 16223 that is one of the HV1 diagnostic positions of the haplogroup; they also included two individuals carrying an L2 sequence that is identical in two Pasiegos and has also been found in the African archipelago of Cabo Verde (Brehm *et al.* 2002). Nonetheless, summing up the percentage of putative African sequences in these samples (2% in pooled Cantabrians and 4% in Pasiegos) it is rather similar to those found in other areas of the Iberian Peninsula (Corte-Real *et al.* 1996; González *et al.* 2003; Larruga *et al.* 2001; Pereira *et al.* 2000; Salas *et al.* 1998). The comparison of Cantabrian samples to adjacent populations (Table 5) reinforces the haplogroup frequency differentiation of Pasiegos and pooled Cantabrians from the rest of populations, including Lebaniegos, by their comparatively high frequencies of V and low frequencies of H. The exclusive presence of U5a representatives in Basques is unique in the Cantabrian edge. This sub-haplogroup seems to be more frequent in southern than northern Europe (Richards *et al.* 1998), being also detected in North Africa (Rando *et al.* 1998). Also striking is the respective presence/absence of the I haplogroup in all Cantabrian samples, and the W and X incidence in Basques and Galicians.

Table 4 HVS I haplotypes and RFLPs in Cantabrians, Lebaniegos and Pasiegos samples

Sequence type HVS-I (16...)	Populations		
	CAN	LEB	PAS
H -7025 AluI			
CRS	9	18	7
093	2	1	
129	3	3	2
145			1
188G	1		
192	1		
220		1	
233			3
248	1	3	
264	1		1
288	1		1
291		4	
304	3		
311	1		
362	3	1	
093 291		1	
093 293	1		
148 233	1		2
193 219	1		
213 362		1	
239 242	1	1	3
269 270		1	
051 162 264			2
111 220C 362		1	
188G 189 311		1	
080 081 189 356	1		
080 081 189 309 356		1	
	31	38	22
V (298) -4577 NlaIII			
298	5		8
221 291		1	
240 298	3	5	2
249 298	2		2
298 311	1		
189 240 298	2		1
213 221 291		2	
	13	8	13
preV-I +00071 SphI +4577 NlaIII -15904 Tru9I			
298	6	1	4
	6	1	4
PreHV (126 362) +14766 MseI +11718 HaeIII +12705 NdeII			
126 355 362	1		
	1	0	0
HV -00071 SphI +4577 NlaIII +11718 HaeIII -14766 MseI -15904 Tru9I			
153 +7025 AluI	1		
298 +7025 AluI	1		
129 298			3
	2	0	3
U* +12308 HinfI			
CRS +7025 AluI +12308 HinfI	1		1
093 189 +7025 AluI +12308 HinfI		1	

Table 4 (continued)

Sequence type HVS-I (16...)	Populations		
	CAN	LEB	PAS
189 274 +7025 <i>AluI</i> +12308 <i>HinfI</i>		1	
278 311 +7025 <i>AluI</i> +12308 <i>HinfI</i>	1		
	2	2	1
U2 (129C 189) +12308 <i>HinfI</i>			
051 129C 189 362	1		
051 129C 189 264 362	1		
	2	0	0
U3 (343) +12308 <i>HinfI</i>			
168 343	2		1
	2	0	1
U4 (356) +12308 <i>HinfI</i>			
356	2	1	1
134 356	1		1
278 356	1		
	4	1	2
U5* (270) +12308 <i>HinfI</i>			
224 270 311 +9052 <i>HaeII</i>			4
224 269 270 311		1	
224 270 311 390 +9052 <i>HaeII</i>	4		6
	4	1	10
U5a1 (192 256 270) +12308 <i>HinfI</i>			
172 192 232 256 270 291	1	1	
	1	1	0
U5a1a (256 270) +12308 <i>HinfI</i>			
189 256 270	1		
	1	0	0
U5b (189 270) +12308 <i>HinfI</i>			
189 270	1		
074 189 270	1		
189 270 311	1		1
189 270 274 311	1		
172 189 192 270 311		1	2
	4	1	3
U6a1 (172 189 219 278) +12308 <i>HinfI</i>			
172 189 219 239 278	1		
	1	0	0
K (224 311) -9052 <i>HaeII</i>			
224 311	1	2	
093 224 311			4
093 224 311 319		2	
	1	4	4
T* (126 294) +4216 <i>NlaIII</i>			
126 294	1		
126 294 296			3
126 172 245 294 296		3	
	1	3	3
T1 (126 163 186 189 294) +4216 <i>NlaIII</i>			
037 126 163 186 189	1		
	1	0	0
T2 (126 294 304) +4216 <i>NlaIII</i>			
126 294 296 304		2	
126 239 294 296 304	1		
126 241 294 296 304	1		2

Table 4 (continued)

Sequence type HVS-I (16...)	Populations		
	CAN	LEB	PAS
T3 (126 292 294) +4216 NlaIII	2	2	2
126 292 294	1		6
	1	0	6
T5 (126 153 294) +4216 NlaIII			
126 153 294	1	1	
126 153 209 294		2	
	1	3	0
J* (069 126) +4216 NlaIII			
069 126	1		
069 126 172		1	
069 126 300	1		
069 126 366	1		
069 126 278 366	1	1	
	4	2	0
J1 (069 126 261) +4216 NlaIII			
069 126 261		1	
	0	1	0
J1b (069 126 145 222 261) +4216 NlaIII			
069 126 145 222 261		1	
	0	1	0
J1b1 (069 126 145 172 222 261) +4216 NlaIII			
069 126 145 172 222 261		1	
	0	1	0
I (129 223) +10032 AluI			
129 145 223 391		1	1
129 223 301 391			1
129 223 311 391	1		
129 145 223 301 391			3
129 172 223 311 391		1	
	1	2	5
M1(129 189 223 249 311) +10394 DdeI +10397 AluI			
129 189 249 311			1
	0	0	1
L2(223 278 390) +3592 HpaI			
223 278 320 390	2		2
	2	0	2
Total sample	88	72	82

Population codes as in Table 2.

Global affinity measures between populations (Table 6) show that mtDNA has more geographical differentiation than Y chromosome polymorphisms. As in the Y-chromosome analysis, pooled Cantabrians and Pasiegos constitute a homogeneous group; however, the Lebaniegos clearly show more affinities to Basques and Galicians than to other Cantabrians. The previously detected similarity between Basques and Galicians (Salas *et al.* 1998) at mtDNA level is also confirmed here. The fact that the same relationships stand using either

F_{ST} comparisons based on haplogroup frequencies, or i comparisons based on haplotypic matches (Table 6), strengthens the validity of these affinities.

MtDNA haplotypic diversity values for each population are depicted on the diagonal of Table 6. Again, the lowest diversity corresponds to the Pasiego isolate.

Discussion

The comparatively low haplotypic diversity found, and the lack of lineages with Near Eastern assignment for

Table 5 MtDNA haplogroup frequencies in Northern Spain

Haplogroup	Populations				
	GAL ¹	CAN	LEB	PAS	BAS
H/HV*/U*/R*	62.9	39.8	55.5	31.8	62.2
pre-*HV	0.7	1.1	0	0	0
V/pre-V-I	5.2	21.6	12.5	20.7	10.2
U2	1.5	2.3	0	0	0.6
U3	0	2.3	0	1.2	0
U4	0.7	4.6	1.4	2.4	0
U5*	0	4.6	1.4	12.2	0.6
U5a*	0	0	0	0	10.8
U5a1*	1.5	1.1	1.4	0	0
U5a1a	0	1.1	0	0	0
U5b	1.5	4.6	1.4	3.7	0.6
U6	2.3	1.1	0	0	0.6
U7	0	0	0	0	0.6
K	5.2	1.1	5.6	4.9	3.4
T*	0	1.1	4.2	3.7	2.4
T1	0.7	1.1	0	0	0
T2	2.3	2.3	2.8	2.4	0.6
T3	0	1.1	0	7.3	0.6
T4	0	0	0	0	0.6
T5	0.7	1.1	4.2	0	0.6
J*	5.2	4.6	2.7	0	1.8
J1*	0	0	1.4	0	0
J1a*	0	0	0	0	0.6
J1b*	0	0	1.4	0	0
J1b1	0.7	0	1.4	0	0
J2	1.5	0	0	0	0
I	0	1.1	2.7	6.1	0
W	2.3	0	0	0	1.2
X	0.7	0	0	0	1.2
M1	0	0	0	1.2	0
N/M*/L3a*	1.5	0	0	0	0.6
L3b	0.7	0	0	0	0
L1b	1.5	0	0	0	0
L2	0.7	2.3	0	2.4	0

¹Galicians, other population abbreviations as in Table 2

Table 6 Linearized F_{ST} between populations based on mtDNA haplogroup frequencies above diagonal. Population mtDNA haplotypic diversities on the diagonal. Identities (i) between populations based on mtDNA haplotype matches below diagonal

	GAL ¹	PAS	LEB	CAN	BAS
GAL	0.60	0.102***	0.004	0.057***	0.008
PAS	0.12	0.38	0.044**	0.004	0.097***
LEB	0.18	0.20	0.53	0.017*	0.008
CAN	0.20	0.52	0.32	0.58	0.055***
BAS	0.23	0.11	0.18	0.17	0.40

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

¹Galicians, other population abbreviations as in Table 2.

Y-chromosome (J-12f2) and mtDNA (J), confirm that the Pasiegos are a rather isolated group. However, this isolation seems to be recent and does not imply strong founder effects, as their striking similarity to the pooled Cantabrians suggests. More surprising is the divergence of the Lebaniego sample, which shows greater affinities to their western neighbours (Asturians and Galicians) than to those of the same community. Nevertheless, it has to be taken into account that nowadays the administrative borders of Cantabria and other northern Spanish provinces do not accurately reflect the historic past of the Cantabrian Cornice (González-Echegaray, 1993). Taken together, Pasiegos and pooled Cantabrians stand out due to a high presence of North African lineages (E-M81 Y-haplotype and, to a lesser degree, U6, M1 and L2 mtDNA lineages). Compared to other North European populations they also share high frequencies of R-SRY10831b Y-chromosome lineage and of mtDNA lineages V and I. For the African input, it is tempting to propose an asymmetric sexual contribution in which males predominate. Although it could be true, this supposition should be regarded with caution, since the majority of North African mtDNA lineages are also present in Europe (Rando *et al.* 1998). There are, currently, two theories to explain this not negligible African input on the Iberian peninsula. For some, it is mainly the result of the historic Islamic occupation (Bosch *et al.* 2001; Pereira *et al.* 2000), whereas others, without totally denying this possibility, favour the bulk of this influence as having prehistoric roots (Gómez-Casado *et al.* 2000; González *et al.* 2003). If the sexual asymmetry could be demonstrated the former hypothesis would be strengthened. In any case, the fact that this African influence similarly affects other Cantabrians and the lack of assigned Near East lineages in Pasiegos, rules out the hypothesis that this isolate was specifically founded by Moorish and Jewish refugees. The second component points to some kind of relationship between Cantabrians and Northern Europeans. It could be explained as a result of the well documented waves of Northern tribes that, throughout the Pyrenees, have repeatedly invaded the Iberian Peninsula since prehistoric times. R-SRY10831b frequencies in Pasiegos (18%) are only surpassed, in a European context, by northeastern populations (Scozzari *et al.* 2001), but direct influences from this area are weakened by the lack in Pasiegos of

N-TatC which is also highly frequent in those populations. On the other hand, the relatively low number of V sequences found precludes any temporal inference of this haplogroup in Cantabria based on a haplotype phylogeny. Relative V matches of Cantabrians to the Iberian Peninsula or Northern Europe are not informative either. Excluding the central motif (16298), there are four different V sequences in Cantabrians (Lebaniegos excluded): 16298, 16311 is a very widespread type; 16240, 16298 has been found in Portugal but also in Northern Europe; for 16249, 16298 and 16189, 16240, 16298 no matches have been found. However, it is noticeable that the frequencies of the ancestral lineages of V (pre-V) in pooled Cantabrians (7%) and Pasiegos (5%) are at the top of their range in Europe, and contrast with the lack of pre-V sequences in the nearby populations such as Basques, Portuguese or French and the low frequencies detected in northern Europe (Torroni *et al.* 2001). These results are congruent with the hypothesis that haplogroup V arose in a south-western European glacial refugium from which a postglacial recolonization of Europe was initiated (Torroni *et al.* 1998; Torroni *et al.* 2001). The Basque country was suggested as a possible candidate focus of this dispersion, but this was refuted on the basis of the lack of haplogroup V representatives in prehistoric Basques (Izagirre *et al.* 1999). In the light of our results, this focus could well be displaced to Cantabria. Nevertheless, a more recent northern invasion bringing V sequences that mixed with a recipient Cantabrian population harbouring mainly pre-V sequences is also a plausible alternative. The most probable hypothesis about the origin of Pasiegos, which also holds for other Cantabrians, is that they are the result of an ancient indigenous substrate more or less mixed with more recent immigrants. The other possibility, that the high frequencies found in Cantabrians for lineages with dual geographic origins have been the result of genetic drift, is weakened because it happens in outbred and inbred samples, and it is recurrent in independent male and female genetic lineages as well as in autosomal markers (Esteban *et al.* 1998; Sánchez-Velasco *et al.* 1999).

Even today, Cantabrians (the Pasiego included, Lebaniegos excluded), at the North of the Iberian Peninsula, seem to be a genetically well differentiated community, as deduced from uniparental and autoso-

mal (Esteban *et al.* 1998; Sánchez-Velasco *et al.* 2003) markers, perhaps to a higher degree than their neighbours, the Basques, who are the best known European isolate on linguistic grounds.

Acknowledgements

This work was supported by grants from the Ministerio de Educación (PB96-1034), the Fundación Marqués de Valdecilla (FMV 16/99) and the Parlamento de Cantabria.

References

- Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
- Bischoff, J., García, J. & Straus, L. (1992) Uranium-series isochron dating at El Castillo Cave (Cantabria, Spain). The “Acheulean”/“Mousterian” question. *J Archaeol Sci* **16**, 49–62.
- Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A. & Bertranpetit, J. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian peninsula. *Am J Hum Genet* **68**, 1019–1029.
- Brehm, A., Pereira, L., Bandeltm J. H., Prata, M. J. & Amorim, A. (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann Hum Genet* **66**, 49–60.
- Corte-Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B. C. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331–350.
- Esteban, E., Dugoujon, J. M., Guitardm E., Senegas, M. T., Manzano, C., de la Rua, C., Valveny, N. & Moral, P. (1998) Genetic diversity in northern Spain (Basque Country and Cantabria): GM and KM variation related to demographic histories. *Eur J Hum Genet* **6**, 315–324.
- Fernández-Acebo V. (1996) *Notas para el conocimiento de la etnogenia y demografía en los Montes de Pas*. Vega de Pas: Asociación de Estudios Pasiegos.
- Gómez-Casado, E., del Moral, P., Martínez-Laso, J., García-Gómez, A., Allende, L., Silvera-Redondo, C., Longas, J., González-Hevilla, M., Kandil, M., Zamora, J. & Arnaiz-Villena, A. (2000) HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens* **55**, 239–249.
- González, A. M., Brehm, A., Pérez, J. A., Maca-Meyer, N., Flores, C. & Cabrera, V. M. (2003) Mitochondrial DNA

- affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol* In press.
- González-Echegararay, J. (1993) *Los cántabros*. Santander: Ediciones de Librería Estudio.
- Helgason, A., Sigurethardóttir, S., Gulcher, J. R., Ward, R. & Stefánsson, K. (2000) mtDNA and the origin of the Icelanders: Deciphering signals of recent population history. *Am J Hum Genet* **66**, 999–1016.
- Izagirre, N. & de la Rúa, C. (1999) A 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.
- Larruga, J. M., Díez, F., Pinto, F. M., Flores, C. & González, A. M. (2001) Mitochondrial DNA characterisation of European isolates: The Maragatos from Spain. *Eur J Hum Genet* **9**, 708–716.
- Pereira, L., Prata, M. J. & Amorim, A. (2000) mtDNA diversity in Portugal: not a genetic edge of European variation. *Ann Hum Genet* **64**, 491–506.
- Pérez-Flores, E. (1998) *Estudio ecológico en la comunidad pasiega* (PhD thesis). Madrid: Universidad Autónoma.
- Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H. J. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* **62**, 531–550.
- Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H. J. & Sykes, B. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* **59**, 185–203.
- Richards, M. B., Macaulay, V. A., Bandelt, H. J. & Sykes, B. C. (1998) Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* **62**, 241–260.
- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C. & Stefanescu, G. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.
- Rudbeck, L. & Dissing, J. (1998) Rapid, simple alkaline extraction of human genomic DNA from whole blood, buccal epithelial cells, semen and forensic stains for PCR. *Biotechniques* **25**, 588–590, 592.
- Salas, A., Comas, D., Lareu, M. V., Bertranpetit, J. & Carracedo, A. (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* **6**, 365–375.
- Sánchez-Velasco, P., Escribano de Diego, J., Paz-Miguel, J. E., Ocejo-Vinyals, G. & Leyva-Cobián, F. (1999) HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations. *Tissue Antigens* **53**, 65–73.
- Sánchez-Velasco, P., Gómez-Casado, E., Martínez-Laso, J., Moscoso, J., Zamora, J., Lowy, E., Silvera, C., Cemborain, A., Leyva-Cobián, F. & Arnaiz-Villena, A. (2003) HLA alleles in North Spain isolated populations: origin of the Basques and the ancient Iberians. *Tissue Antigens* (in press).
- Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin ver. 2000: A software for population genetic data analysis*. Switzerland, Genetics and Biometry Laboratory, University of Geneva.
- Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Laitini, V., Varesi, L., Memmi, M. M., Romano, V., De Leo, G., Gennarelli, M., Jaruzelska, J., Villems, R., Parik, J., Macaulay, V. & Torrioni, A. (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* **62**, 871–884.
- Sun, C., Skaletsky, H., Rozen, S., Gromoll, J., Nieschlag, E., Oates, R. & Page, D. C. (2000) Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum Mol Genet* **9**, 2291–2296.
- Tax-Freeman, S. (1979) *The Pasiegos. Spaniards in no man's land*. Chicago: The University of Chicago Press.
- Torrioni, A., Bandelt, H. J., D' Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.L., Bonne-Tamir, B. & Scozzari R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* **62**, 1137–1152.
- Torrioni, A., Bandelt, H. J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martínez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M.L., Huoponen, K., Laitinen, V., Koivumaki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A. S., Semino, O. & Scozzari, R. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69**, 844–852.
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358–361.
- Y Chromosome Consortium. (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* **12**, 339–348.

Received: 10 October 2002

Accepted: 31 March 2003