Ana M. Pérez-Miranda · Miguel A. Alfonso-Sánchez
Arif Kalantar · Susana García-Obregón
Marian M. de Pancorbo · José A. Peña
Rene J. Herrera

Microsatellite data support subpopulation structuring among Basques

Received: 13 April 2005 / Accepted: 6 June 2005 / Published online: 30 August 2005
© The Japan Society of Human Genetics and Springer-Verlag 2005

Abstract Genomic diversity based on 13 short tandem repeat (STR) loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, TH01, TPOX, and CSF1PO) is reported for the first time in Basques from the provinces of Guipúzcoa and Navarre (Spain). STR data from previous studies on Basques from Alava and Vizcaya provinces were also examined using hierarchical analysis of molecular variance (AMOVA) and genetic admixture estimations to ascertain whether the Basques are genetically heterogeneous. To assess the genetic position of Basques in a broader geographic context, we conducted phylogenetic analyses based on F_{ST} genetic distances [neighbor-joining trees and multidimensional scaling (MDS)] using data compiled in previous publications. The genetic profile of the Basque groups revealed distinctive regional partitioning of short tandem repeat (STR)

diversity. Consistent with the above, native Basques clearly segregated from other populations from Europe (including Spain), North Africa, and the Middle East. The main line of genetic discontinuity inferred from the spatial variability of the microsatellite diversity in Basques significantly overlapped the geographic distribution of the Basque language. The genetic heterogeneity among native Basque groups correlates with the peculiar geography of peopling and marital structure in rural Basque zones and with language boundaries resulting from the uneven impact of Romance languages in the different Basque territories.

Keywords Short tandem repeats · Microsatellite diversity · Linguistic barrier · Population genetics · Genetic heterogeneity · Basques

Ana M. Pérez-Miranda and Miguel A. Alfonso-Sánchez contributed equally to this work.

A.M. Pérez-Miranda · A. Kalantar · R.J. Herrera (✉)
Molecular Biology and Human Diversity Laboratory,
Department of Biological Sciences, Florida International
University, Miami, FL 33199, USA
E-mail: herrerar@fiu.edu
Tel.: +1-305-3481258
Fax: +1-305-3481259

A.M. Pérez-Miranda · M.A. Alfonso-Sánchez
S. García-Obregón · J. A. Peña
Dpto de Genética y Antropología Física,
Universidad del País Vasco, Apartado 644,
Vizcaya, 48080 Bilbao, Spain

A. Kalantar
General Department of Forensic Services,
Biology and DNA Section,
Dubai Police H.Q.,
United Arab Emirates

M.M. de Pancorbo
Dpto de Zoología y Dinámica Celular Animal,
Facultad de Farmacia,
Universidad del País Vasco,
Vizcaya, 48940 Leioa, Spain

Introduction

With the development of rapid screening techniques using polymerase chain reaction (PCR) amplification, the possibilities of identifying highly variable DNA markers and, consequently, of studying genetic structure and microdifferentiation processes in human populations accurately, have increased remarkably. Among the molecular markers that can be easily genotyped and scored by PCR-based techniques, short tandem repeats (STRs) stand out as being abundant and widespread throughout the human genome. Microsatellites or STRs are short sequences of DNA with units 2–6 bp in length (Hearne et al. 1992), which are repeated numerous times in a head–tail manner and ubiquitously distributed in the genome. STRs have been previously employed in the elucidation of human population history (Jorde et al. 1997; Rowold and Herrera 2003; Zhivotovsky et al. 2004) and subpopulation structure (Rowold and Herrera 2005).

The genetic uniqueness of the Basques has long been recognized (Mourant 1947). According to their anatomical, archaeological, linguistic, and genetic singular-

ities, the Basques have been considered among the most ancient inhabitants of Europe as well as one of the oldest human isolates (Cavalli-Sforza et al. 1994). For this reason, the autochthonous groups of the traditional Basque territories have been the subject of a great number of biological studies and have been characterized on the basis of their genetic peculiarities and geographic diversity (reviewed in Calderón et al. 1998). However, the issues concerning the origins of this interesting group of humans remain a subject of dispute for scientists. Some authors suggest an upper Paleolithic origin for the Basque people based on findings of population genetic studies using classical markers such as blood groups, serum proteins and enzymes (Calafell and Bertranpetit 1994), minisatellites (Alonso and Armour 1998), Y-chromosomal single nucleotide polymorphisms (SNPs) (Lucotte and Hazout 1996), and mitochondrial DNA (Bertranpetit et al. 1995). Conversely, data on immunoglobulin allotypes support a more recent Neolithic origin (Calderón et al. 1998).

As for their genetic structure, the existence of a certain degree of genetic heterogeneity within the Basques has been suggested since the first population-based genetic analyses were performed (Goedde et al. 1972, 1973). On the one hand, some authors have claimed a lack of genetic substructure among the Basques based on investigations using both classical markers (Calafell and Bertranpetit 1994) and HLA genetic markers (Comas et al. 1998). Subsequently, a sizable number of genetic studies have provided a conflicting set of data regarding genetic variability of autochthonous Basque groups. Recent studies designed to assess the geographic patterning of the Basques indicate that their genetic diversity is spatially structured (Aguirre et al. 1991; Manzano et al. 2002). This contention is starting to be confirmed as well by DNA molecular markers (Brown

et al. 2000; Pancorbo et al. 2001; Iriondo et al. 2003; Pérez-Miranda et al. 2003). Some results point to Guipúzcoa as being the Basque province with the most genetic distinctiveness whereas the highest levels of genetic affinity with non-Basque surrounding populations have been reported for the province of Alava (Calderón et al. 1998; Pérez-Miranda et al. 2003).

The lack of consensus concerning genetic diversity among the different Basque autochthonous groups prompted us to study the Basques according to STR polymorphisms. Thus, in the present study, we analyzed a set of 13 STR loci of 4 bp motifs (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, TH01, TPOX, and CSF1PO) with the aim of characterizing genetically the autochthonous Basque groups settled in the provinces of Guipúzcoa and Navarre (northern Spain). These STRs constitute the core of PCR-based genetic markers in the US-combined DNA index system (CODIS). In addition, STR allelic frequency data previously reported by our research team from two Basque groups from the provinces of Alava and Vizcaya were included in our analyses to augment the geographical scope of our study (Pérez-Miranda et al. 2005a, 2005b). Finally, for the purposes of assessing population affinities and phylogenetic relationships with other human groups, European (including Spain), North African, and Middle Eastern populations were jointly analyzed.

Materials and methods

Populations studied

The Basque Country administratively includes several French and Spanish provinces in which the Basque

Fig. 1 Geographic location of the populations included in a phylogenetic analysis based on 13 short tandem repeat (STR) loci. The groups examined are Guipúzcoa (1) and North Navarre (2). Other Basque groups: Alava (3), Vizcaya (4), and Residents in the Basque Country (5). Other Spanish collections: Northern Spain (6), Andalusia (7), Extremadura (8), and Canary Islands (9). Other European samples: Portugal (10), Italy (11), Turkey (12), Switzerland (13), Poland (14). North African populations: Morocco (15) and Egypt (16). Middle Eastern populations: Syria (17) and United Arab Emirates (18)

language (*Euskera*) is still spoken (to different degrees) as the mother language. In Spain, the Basque territory lies at the northern region of the Iberian Peninsula and is formed by the autonomous community of the Basque Country which includes the provinces of Alava, Vizcaya, and Guipúzcoa as well as the Chartered Community of Navarre (province of Navarre).

Guipúzcoa is the only Basque “historical territory” that is completely surrounded by other provinces where Basque speakers are native (Fig. 1). The Basque area was among the pioneering Spanish regions embracing industrialization. The process of industrialization in the Spanish Basque Country started by the middle of the nineteenth century, mainly in Vizcaya and Guipúzcoa. Consequently, specifically in Guipúzcoa, over the 1860–1900 period, the population density increased by about 20%. These immigrants came mostly from the bordering Spanish provinces. The development of the Basque industry reached its height by the beginning of 1950 and it coincided with the industrial revolution in Spain. It is estimated that around 30% of the current population of Guipúzcoa is the result of the continuous, large-scale immigration that took place from 1950 to 1980 when the Basque territory was immersed in a prosperous industrialization process. However, these spectacular demographic changes promoted by industrialization occurred mainly in zones close to the industrial centers.

The province of Navarre has been historically populated by autochthonous Basques. It has been argued, however, that much of the present-day Navarrese territory cannot really be considered anthropologically Basque; rather, native Basques there seem to be mostly confined to its northernmost part (Calderón et al. 1998). In contrast with the early modernization of Guipúzcoa, Navarre’s industrialization process did not come until the 1960s, so the demographic size of Navarre has not increased notably and the population contribution of this region in the context of overall Basque demography has dropped considerably over the last 100 years.

Samples and STR genotyping

Whole blood samples were collected in EDTA vacutainer tubes by venipuncture from unrelated healthy autochthonous individuals from Guipúzcoa ($n=102$) and from the north of Navarre ($n=112$). Basque ancestry was ascertained for three generations back in order to define autochthony. Adherence to ethical guidelines was followed as stipulated by each of the institutions involved in the study.

Genomic DNA was extracted by the standard phenol–chloroform procedure (Maniatis et al. 1982). For each sample, 13 loci were amplified simultaneously using AmpF/STR Profiler Plus and AmpF/STR COfiler PCR Amplification Kits (Applied Biosystems, Foster City, CA, USA) at the D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D16S539, TH01, TPOX, CSF1PO, and D7S820 STR loci. PCR amplifications were performed as described in the kit user manual using the recommended DNA amount (1.0–2.5 ng) in a final PCR volume of 12 μ l. DNA was amplified in a GeneAmp PCR System 9600 thermal cycler (Perkin-Elmer Applied Biosystems, Foster City, USA). Amplified STR fragments were analyzed with an ABI PRISM 377 DNA Sequencer (Perkin-Elmer Applied Biosystems). An internal size standard (GeneScan 500 ROX, Perkin-Elmer Applied Biosystems) was included. Genotyping of each sample was made using Genotyper 3.7 NT and GeneScan 3.7 software by comparison with supplied allelic ladders. Allelic designations followed the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics (DNA recommendations, 1994).

Phylogenetic and statistical analyses

Allelic frequencies for the 13 STR loci in the Basque groups from Guipúzcoa and Navarre were estimated

Table 1 Populations included in a phylogenetic analysis based on 13 short tandem repeat (STR) loci

	Populations	Label	References
1	Guipúzcoa	GUIP	Present study
2	North Navarre	NNAV	Present study
3	Alava	ALAV	Pérez-Miranda et al. (2005a)
4	Vizcaya	VIZC	Pérez-Miranda et al. (2005b)
5	Resident Basques	RBAS	García et al. (2003)
6	Northeast Spain	NSPA	Paredes et al. (2003)
7	Andalucía	ANDL	Sanz et al. (2001)
8	Extremadura	EXTR	García-Hirschfeld et al. (2003)
9	Canary Island	CANR	http://www.unniduesseldorf.de
10	Portugal	PORT	http://www.unniduesseldorf.de
11	Italy	ITAL	Garofano et al. (1998)
12	Turkey	TURK	Akbasak et al. (2001)
13	Switzerland	SWIT	http://www.unniduesseldorf.de
14	Poland	PLND	Pawlowski and Maciejewska (2000)
15	Morocco	MORC	Abdin et al. (2003)
16	Egypt	EGYP	http://www.unniduesseldorf.de
17	Syria	SYRI	Abdin et al. (2003)
18	United Arabs Emirates	UAE	Alshamali et al. (2003)

Table 2 Allelic frequencies at 13 short tandem repeat (STR) loci in autochthonous Basques from Guipúzcoa province (Spain)

Number	D3S1358 (n = 102)	VWA (n = 101)	FGA (n = 100)	D8S1179 (n = 94)	D21S11 (n = 100)	D18S51 (n = 93)	D5S818 (n = 93)	D13S317 (n = 101)	D7S820 (n = 102)	D16S539 (n = 100)	TH01 (n = 102)	TPOX (n = 102)	CSF1PO (n = 102)
5	—	—	—	—	—	—	—	—	—	—	0.0245	—	—
6	—	—	—	—	—	—	—	—	—	—	0.1961	0.0049	—
7	—	—	—	—	—	—	—	—	0.0196	—	0.0882	0.0441	—
8	—	—	—	0.0106	—	—	0.0054	0.2970	0.1961	0.0050	0.1765	0.4951	—
9	—	—	—	0.0372	—	—	0.0215	0.0198	0.0588	0.0850	0.2500	0.0931	—
9.3	—	—	—	—	—	—	—	—	—	—	0.2647	—	—
10	—	—	—	0.0851	—	0.0161	0.1452	0.0594	0.3186	0.0500	—	0.0882	0.3480
11	—	—	—	0.0426	—	0.0054	0.3763	0.3069	0.2304	0.2500	—	0.2451	0.2696
12	0.0049	0.0050	—	0.1064	—	0.1882	0.2957	0.1980	0.1177	0.3500	—	0.0294	0.3284
13	—	—	—	0.2819	—	0.1613	0.1398	0.0941	0.0294	0.2300	—	—	0.0490
14	0.1471	0.0842	—	0.2713	—	0.1075	0.0161	0.0248	0.0245	0.0300	—	—	—
15	0.3235	0.0990	—	0.1436	—	0.1452	—	—	0.0049	—	—	—	0.0049
16	0.1471	0.2277	—	0.0213	—	0.0699	—	—	—	—	—	—	—
17	0.1275	0.3168	—	—	—	0.1774	—	—	—	—	—	—	—
18	0.2353	0.1832	0.0350	—	—	0.0484	—	—	—	—	—	—	—
19	0.0147	0.0792	0.1050	—	—	0.0484	—	—	—	—	—	—	—
20	—	0.0050	0.1850	—	—	0.0054	—	—	—	—	—	—	—
21	—	—	0.1400	—	—	0.0215	—	—	—	—	—	—	—
22	—	—	0.1050	—	—	0.0054	—	—	—	—	—	—	—
23	—	—	0.1600	—	—	—	—	—	—	—	—	—	—
24	—	—	0.1500	—	—	—	—	—	—	—	—	—	—
25	—	—	0.0650	—	—	—	—	—	—	—	—	—	—
26	—	—	0.0200	—	—	—	—	—	—	—	—	—	—
26.5	—	—	0.0050	—	—	—	—	—	—	—	—	—	—
27	—	—	—	—	0.0050	—	—	—	—	—	—	—	—
28	—	—	0.0300	—	0.0600	—	—	—	—	—	—	—	—
29	—	—	—	—	0.1800	—	—	—	—	—	—	—	—
30	—	—	—	—	0.3300	—	—	—	—	—	—	—	—
30.2	—	—	—	—	0.0400	—	—	—	—	—	—	—	—
31	—	—	—	—	0.0700	—	—	—	—	—	—	—	—
31.2	—	—	—	—	0.0700	—	—	—	—	—	—	—	—
32.2	—	—	—	—	0.1700	—	—	—	—	—	—	—	—
33.2	—	—	—	—	0.0650	—	—	—	—	—	—	—	—
34.2	—	—	—	—	0.0100	—	—	—	—	—	—	—	—

Table 3 Allelic frequencies at 13 short tandem repeat (STR) loci in autochthonous Basques from Navarre province (Spain)

Number	D3S1358 (<i>n</i> = 112)	VWA (<i>n</i> = 112)	FGA (<i>n</i> = 112)	D8S1179 (<i>n</i> = 112)	D21S11 (<i>n</i> = 112)	D18S51 (<i>n</i> = 112)	D5S818 (<i>n</i> = 112)	D13S317 (<i>n</i> = 112)	D7S820 (<i>n</i> = 112)	D16S539 (<i>n</i> = 112)	TH01 (<i>n</i> = 109)	TPOX (<i>n</i> = 112)	CSF1PO (<i>n</i> = 112)
5	—	—	—	—	—	—	—	—	—	—	0.0275	—	—
6	—	—	—	—	—	—	—	—	—	—	0.2615	—	—
7	—	—	—	—	—	—	0.0045	—	0.0223	—	0.1330	0.0089	—
8	—	—	—	0.0134	—	—	—	0.2143	0.1250	—	0.1330	0.5848	—
9	—	—	—	0.0045	—	—	0.0179	0.0446	0.1071	0.1027	0.1651	0.0313	0.0313
9.3	—	—	—	—	—	—	—	—	—	—	0.2706	—	—
10	—	—	—	0.0670	—	0.0134	0.0670	0.0759	0.3705	0.0089	0.0092	0.0714	0.3304
11	—	—	—	0.0446	—	0.0089	0.3705	0.2857	0.1920	0.2768	—	0.2768	0.3438
12	—	—	—	0.1027	—	0.2054	0.3527	0.2366	0.1250	0.3036	—	0.0268	0.2500
13	0.0045	0.0045	—	0.2991	—	0.0580	0.1830	0.0670	0.0402	0.2545	—	—	0.0402
13.2	—	—	—	—	—	—	0.0045	—	—	—	—	—	—
14	0.1116	0.1250	—	0.2232	—	0.2009	—	0.0714	0.0179	0.0446	—	—	0.0045
15	0.2411	0.1429	—	0.2098	—	0.1786	—	0.0045	—	0.0089	—	—	—
16	0.2411	0.1607	—	0.0268	—	0.0848	—	—	—	—	—	—	—
17	0.1339	0.3125	—	0.0045	—	0.1071	—	—	—	—	—	—	—
18	0.2500	0.2009	0.0670	0.0045	—	0.0357	—	—	—	—	—	—	—
19	0.0179	0.0402	0.1429	—	—	0.0491	—	—	—	—	—	—	—
20	—	0.0089	0.1250	—	—	0.0446	—	—	—	—	—	—	—
21	—	0.0045	0.1920	—	—	0.0134	—	—	—	—	—	—	—
22	—	—	0.0982	—	—	—	—	—	—	—	—	—	—
23	—	—	0.1295	—	—	—	—	—	—	—	—	—	—
24	—	—	0.1205	—	—	—	—	—	—	—	—	—	—
24.2	—	—	—	—	0.0045	—	—	—	—	—	—	—	—
25	—	—	0.0938	—	0.0045	—	—	—	—	—	—	—	—
26	—	—	0.0313	—	—	—	—	—	—	—	—	—	—
27	—	—	—	—	—	—	—	—	—	—	—	—	—
28	—	—	—	—	0.0357	—	—	—	—	—	—	—	—
29	—	—	—	—	0.0982	—	—	—	—	—	—	—	—
29.2	—	—	—	—	0.2143	—	—	—	—	—	—	—	—
30	—	—	—	—	0.0045	—	—	—	—	—	—	—	—
30.2	—	—	—	—	0.3036	—	—	—	—	—	—	—	—
31	—	—	—	—	0.0313	—	—	—	—	—	—	—	—
31.2	—	—	—	—	0.0491	—	—	—	—	—	—	—	—
32	—	—	—	—	0.0446	—	—	—	—	—	—	—	—
32.2	—	—	—	—	0.1250	—	—	—	—	—	—	—	—
33	—	—	—	—	0.0045	—	—	—	—	—	—	—	—
33.2	—	—	—	—	0.0625	—	—	—	—	—	—	—	—
34.2	—	—	—	—	0.0179	—	—	—	—	—	—	—	—

by direct counting. To test for Hardy–Weinberg equilibrium (HWE) expectations, a Fisher's exact probability test was conducted to estimate P values (Guo and Thompson 1992) using the Arlequin Version 2.000 software (Schneider et al. 2000). Several useful parameters in legal medicine were also calculated for these two autochthonous Basque collections, including polymorphic information content (PIC) (Smouse and Chakraborty 1986) and power of discrimination (PD) (Guo and Thompson 1992). To ascertain phylogenetic relationships based on the allelic frequencies of these STR markers, data compiled from previous studies were used to create a genetic database of European, North African, and Middle Eastern populations (see Fig. 1 and Table 1). Genetic information of these databases was used to compute F_{ST} unbiased genetic distances (Reynolds et al. 1983) between all pairs of populations. From the resultant F_{ST} genetic distance matrix, phylogenetic trees based on the Neighbor-Joining (NJ) method (Saitou and Nei 1987) were constructed using the Phylip Version 3.2 program (Felsenstein 1989). The reliability of the dendrogram was evaluated by bootstrap resampling (Felsenstein 1985). Genetic structuring among various population clusters defined according to geographic criteria was examined through hierarchical analysis of molecular variance (AMOVA) (Excoffier et al. 1992) using the Arlequin program. In this statistical test, a permutation procedure is employed to assess the significance of the F_{SC} and F_{CT} fixation values. These indices reflect the relative contribution of genetic variation among populations within groups and among groups, respectively. In order to represent the F_{ST} genetic distance matrix for the 18 populations examined, a two-dimensional genetic map based on nonmetric multidimensional scaling (MDS) analysis (Kruskal 1964) was generated using the SPSS statistical package. In addition, we used the computer program Structure (Pritchard et al. 2000) to attempt to identify clusters of genetically similar

individuals from multilocus genotypes. The admixture proportions of the different Basque groups included in the study were estimated by mean of the weighted least squares method (Long et al. 1991) mathematically expressed as $p_{ih} = \sum_{j=1}^J p_{ij} \cdot \mu_j$, where p_{ih} is the frequency of the i th allele in the hybrid population, p_{ij} denotes the frequency of the i th allele in the j th reference population ($j=1, J$), μ_j is the proportionate contribution of the j th reference gene pool to the hybrid population and $\sum_{j=1}^J \mu_j = 1$.

Results

STR diversity in Guipúzcoa and Northern Navarre

Tables 2 and 3 provide the allelic frequencies of the 13 STR loci for the autochthonous Basque groups from Guipúzcoa (GUIP) and Northern Navarre (NNAU), respectively. Some alleles commonly detected in STR analyses of worldwide populations could not be identified in the Basque collections under study. This includes alleles 10 of TH01, 12 of TPOX, and 24 of FGA in natives of Guipúzcoa and alleles 32 of D21S11 and 8 of CSF1PO in Northern Navarrese. Similarly, alleles 27 of FGA and 11.2 of D5S818 do not appear in any of the Basque groups analyzed herein.

A number of genetic and forensic parameters of interest were estimated from the STR allelic frequencies and are summarized in Tables 4 (GUIP) and 5 (NNAU). To test for heterozygote deficit, a Fisher's exact probability test was conducted to estimate the P value by the Markov chain Monte Carlo (MCMC) method. HWE expectations were tested for all possible locus-population combinations. No significant departure from HWE expectations was detected suggesting genetic equilibrium for all loci in both GUIP and NNAU samples. Similar results were obtained when HWE was tested through the likelihood ratio test (G test).

Table 4 Statistical parameters of genetic and forensic interest based on 13 short tandem repeat (STR) loci in autochthonous Basques from Guipúzcoa province (Spain). H_o observed heterozygosity; H_e expected heterozygosity; P value HWE, Fisher's ex-

act probability test, G^2 ; P HWE, statistic (G^2) and significance level (P) of the likelihood ratio test (G test); GD gene diversity; PIC polymorphism information content; PD power of discrimination

Locus	Alleles	H_o	H_e	P value	G^2	P	GD	PIC	PD^a
D3S1358	7	0.7843	0.7891	0.5369	14.6649	0.8394	0.7841	0.7404	0.9186
VWA	8	0.7327	0.8077	0.1880	29.6670	0.3793	0.7949	0.7614	0.9267
FGA	11	0.8700	0.8762	0.4556	53.4002	0.5360	0.8736	0.8572	0.9688
D8S1179	9	0.9255	0.8083	0.0106	49.6800	0.0642	0.8083	0.7676	0.9355
D21S11	10	0.7800	0.8145	0.5160	35.8001	0.8348	0.8145	0.7896	0.9420
D18S51	13	0.8495	0.8717	0.0902	67.0832	0.8063	0.8688	0.8456	0.9666
D5S818	7	0.7527	0.7815	0.0167	28.6892	0.1217	0.7335	0.6882	0.8823
D13S317	7	0.7822	0.7743	0.1369	26.5773	0.1853	0.7688	0.7280	0.9078
D7S820	9	0.7549	0.7917	0.5622	25.5483	0.9024	0.7917	0.7567	0.9247
D16S539	7	0.7400	0.7552	0.1766	21.7363	0.4148	0.7552	0.7410	0.8982
TH01	6	0.7941	0.7933	0.2912	19.2789	0.2014	0.7933	0.7511	0.9227
TPOX	7	0.6373	0.7730	0.3658	20.6299	0.4817	0.6788	0.6491	0.8532
CSF1PO	5	0.6373	0.6993	0.1134	14.8891	0.1362	0.6993	0.6322	0.8466

^aCombined $PD = 0.999999999999970$

Table 5 Statistical parameters of genetic and forensic interest based on 13 short tandem repeat (STR) loci in autochthonous Basques from Navarre province (Spain). *Ho* observed heterozygosity; *He* expected heterozygosity; *P* value HWE, Fisher's exact

probability test, G^2 ; *P* HWE, statistic (G^2) and significance level (*P*) of the likelihood ratio test (*G* test); *GD* gene diversity; *PIC* polymorphism information content; *PD* power of discrimination

Locus	Alleles	<i>Ho</i>	<i>He</i>	<i>P</i> value	G^2	<i>P</i>	<i>GD</i>	<i>PIC</i>	<i>PD</i> ^a
D3S1358	7	0.8304	0.7941	0.6265	16.5877	0.7358	0.7941	0.7565	0.9234
VWA	9	0.7679	0.8020	0.3568	29.1449	0.7840	0.8020	0.7711	0.9312
FGA	9	0.8929	0.8807	0.8850	32.9473	0.6146	0.8758	0.8582	0.9699
D8S1179	11	0.7679	0.8023	0.3739	39.2426	0.9464	0.8023	0.7683	0.9315
D21S11	14	0.8482	0.8364	0.4955	59.4494	0.9957	0.8294	0.7759	0.9502
D18S51	12	0.8750	0.8680	0.4017	66.8426	0.4479	0.8613	0.8401	0.9640
D5S818	7	0.7589	0.7031	0.2772	18.5973	0.6110	0.7031	0.6454	0.8554
D13S317	8	0.7679	0.8027	0.8098	17.3395	0.9416	0.8027	0.7707	0.9313
D7S820	8	0.8036	0.7842	0.5011	27.4519	0.4938	0.7842	0.7551	0.9246
D16S539	7	0.6696	0.7572	0.0987	25.9880	0.2069	0.7572	0.7118	0.8974
TH01	7	0.8349	0.7986	0.1211	28.9224	0.1159	0.7985	0.7630	0.9273
TPOX	6	0.5893	0.5771	0.4899	13.2443	0.5834	0.5771	0.5109	0.7608
CSF1PO	6	0.6696	0.7108	0.1477	17.4050	0.2952	0.7108	0.6519	0.8588

^aCombined *PD* = 0.9999999999999996

The observed heterozygosity (*Ho*) in Guipúzcoa ranges from 0.6373 (TPOX and CSF1PO loci) to 0.9255 (D8S1179 locus) whereas in Northern Navarre, *Ho* oscillates between 0.5893 in TPOX and 0.8929 in FGA (see Tables 4 and 5). The combined *PD* value is markedly high in these two Basque groups (GUIP: 0.9999999999999997; NNAV: 0.9999999999999996). As expected, the most polymorphic STR loci are also the most discriminating loci in both Basque collections. This is the case of FGA (GUIP: 96.9%; NNAV: 97.0%), D18S51 (GUIP: 96.7%; NNAV: 96.4%), and D21S11

(GUIP: 94.2%; NNAV: 95.0%) loci, all of which stand out as having the highest observed *Ho* values.

Phylogenetic analyses and genetic structure based on STR diversity patterns

In order to assess the genetic relationships of autochthonous Basques in a broader geographic context and to generate a more complete picture of STR variation, we conducted phylogenetic analyses using additional data compiled from previous studies of Spanish, European, North African, and Middle Eastern populations (see Table 1). With this aim, F_{ST} unbiased genetic distances based on allelic frequencies of the 13 STR loci examined were computed between all pairs of populations. Based on these data, phylogenetic trees using the NJ method were constructed to reveal patterns of geographic associations and population affinities. Although a tree representation has some drawbacks when dealing with populations, it may be useful to recognize clusters of populations with statistical support as given by bootstrap values.

Figure 2 depicts the phylogenetic relationships inferred from STR diversity in the populations examined. In the consensus NJ tree generated, a certain geographic structuring is apparent since, for the most part, the main clusters represent distinct geographic regions. However, the most obvious aspect of this NJ tree is the conspicuous and marked separation of the four autochthonous Basque populations studied (GUIP, NNAV, VIZC, and ALAV) from the remaining populations (including residents in the Basque Country), regardless of their geographic origins. The branch node discriminating the “Basque cluster” shows strong bootstrap support after 1,000 iterations (100%), indicating the high robustness of the topology. Within the Basque cluster, the position of the distinct Basque groups indicates both a high genetic affinity between

Fig. 2 Neighbor-joining (NJ) tree constructed from Reynold's F_{ST} unbiased genetic distances based on the allelic frequencies of 13 short tandem repeat (STR) loci in 18 populations from Europe, North Africa, and the Middle East. Figures in tree nodes are percentage bootstrap values estimated from 1,000 reiterations. Population codes are shown in Table 1

Guipúzcoa and Northern Navarre and the relatively greater genetic distance of the Alava group. These results are in agreement with findings of previous studies on the genetic heterogeneity of Basques where variable levels of genetic substructuring have been reported (Pancorbo et al. 2001; Manzano et al. 2002; Iriondo et al. 2003; Pérez-Miranda et al. 2003).

In addition to the Basque cluster, two other major groupings can be observed in the NJ tree. The biggest of them is exclusively formed by European populations (including Spanish) whereas in the third group, North African (EGYP, MORC) and Middle Eastern (SYRI, UAE) populations segregate together. A more in-depth analysis of the topology of the NJ tree reveals that the residents of the Basque Country (RBAS) occupy an intermediate position between the cluster of autochthonous Basque groups and the remaining Spanish populations (EXTR, CANR, ANDL, and NSPA), as expected, based on the putative mixed nature of its gene pool. An intermediate location is also observed in the case of Turkey (TURK), which segregates between Middle Eastern and European populations.

To further examine how the observed genetic heterogeneity is structured among the Basques, the sample sets were analyzed using AMOVA. The overall estimated F_{ST} was 0.0053 ($P < 0.0001$) indicating a statistically significant STR interpopulation diversity throughout the sampled area. Upon assignment of the populations within the three broad geographic regions as observed in the NJ tree (Basques, Europe, and North Africa plus the Middle East), we obtained an F_{CT} of 0.0045 ($P < 0.05$) and an F_{SC} of 0.0037 ($P < 0.0001$), which suggests significant geographical substructuring involving both interregional and intraregional heterogeneity, respectively.

Additional AMOVA analyses using different hierarchical structures (established according to geography) were performed to obtain maximum genetic variance among groups (F_{CT}) and minimum genetic variance among populations within groups (F_{SC}), which guarantees the statistical consistence of a genetic classification. Of all possible combinations, the hierarchical classification that best fits this criterion was that segregating the whole set of populations into four groups: autochthonous Basques (GUIP, NNAV, VIZC, and ALAV), Spain (RBAS, NSPA, ANDL, EXTR, and CANR), Europe (PORT, ITAL, SWIT, and PLND) and North Africa/Middle East (MORC, EGYP, TURK, SYRI, and UAE). In this case, the corresponding values of the fixation indices were $F_{CT} = 0.0075$ ($P < 0.001$) and $F_{SC} = 0.0011$ ($P < 0.0001$), indicating statistically significant intergroup and intragroup genetic structuring, respectively. It must be emphasized that in spite of the geographical proximity of Portugal (PORT) with the Spanish populations, AMOVA results deteriorated when a regional classification including a group of Iberian populations (Spain/PORT) was employed. A similar situation was found when we performed an AMOVA with the same above-mentioned four groups (Basques, Spain, Europe, and North Africa/Middle East) but this time including Turkey (TURK) within the European group. As far as the Basque groups are concerned (GUIP, NNAV, VIZC, and ALAV), the AMOVA results revealed the existence of statistically significant genetic heterogeneity among these autochthonous collections ($F_{ST} = 0.0015$; $P = 0.0052$).

Figure 3 shows the two-dimensional genetic plot resulting from nonmetric MDS analysis applied on Reynold's F_{ST} genetic distance matrix. The genetic topology is highly robust from the statistical viewpoint,

Fig. 3 Two-dimensional genetic map resulting from nonmetric multidimensional scaling (MDS) applied on Reynold's F_{ST} unbiased genetic distances for 13 short tandem repeat (STR) loci in 18 populations from Europe, North Africa, and the Middle East. The total variance accounted for the eigenvectorial reduction is 93.6%, and the coefficient of stress is 0.1506. Population codes are shown in Table 1

as the vectorial reduction accounts for 93.6% of the total variance. Consistent with the NJ tree and the AMOVA data, three distinct groups are clearly discriminated in the MDS representation. The bulk of European populations (excluding the autochthonous Basques) plotted around the centroid of the distribution as a well-defined cluster. In agreement with the dendrogram, the Spanish groups (NSPA, RBAS, ANDL, CANR, and EXTR) overlap with the remaining of the European populations (PORT, ITAL, SWIT, and PLND). The collection of RBAS plotted in the core of the European cluster and close to the sample of North Spain, as expected, according to their geographical origins. On the other hand, North African and Middle Eastern populations segregated more dispersedly although all of them form a cluster concentrated in the quadrant delimited by the positive segment of both dimensions 1 and 2. In this cluster, the position of Turkey (TURK) is probably the consequence of sharing geographical proximity, historical relations, and common sociodemographic features with Europe, on one hand, and its relationship with Arabic populations from North Africa and the Middle East, on the other.

Regarding the autochthonous Basque groups, most notable was the remote position of Alava with respect to the remaining Basque groups (GUIP, NNAV, and VIZC). As a result of partitioning along axis 2 of the two-dimensional representation, ALAV plotted in the positive upper quadrant, away from GUIP, NNAV and VIZC which segregate in the lower negative quadrant. Of the latter group, GUIP was the Basque autochthonous group segregating more distantly from all other populations. Finally, it should be noted that both the Basque cluster and the group formed by North African and Middle Eastern populations (including Turkey) stand out as plotting in remote positions with respect to the centroid of the two-dimensional map. They occupy extreme and clearly differentiated positions in the distribution. When the Pritchard test was performed on the four Basque groups, it was unable to identify a significant subdivision. It is likely that the number of loci employed in the present study is insufficient to detect subpopulation structure with the Pritchard test.

Admixture coefficients for each Basque group were estimated using the weighted least squares method (Table 6.) Since linguistic, genetic, and sociocultural studies point to Guipúzcoa as the most autochthonous region of the Basque country (Manzano et al. 1996; Calderón et al. 1998), the degree of admixture in the

ALAV, NNAV, VIZC, and RBAS groups was estimated using Guipúzcoa and a Spanish population (ANDL, EXTR, and NSPA) as reference groups. As expected, the genetic pool of RBAS exhibits the minimum proportion of Basque (GUIP) genes (0.256) and the maximum proportion of Spanish genes (0.744). Among the native Basque collections, ALAV (0.348) possesses the lowest proportion of GUIP contribution is ALAV (0.348). The genetic pools of NNAV and VIZC have the least Spanish component, both exhibiting a GUIP component of above 40% (42.0% and 45.6%, respectively). These findings are consistent with the NJ and AMOVA analyses.

Discussion

The allelic frequencies of 13 STR loci in two autochthonous Basque groups from Guipúzcoa and Navarre are reported for the first time. Also, data on the same STR markers previously obtained by our research group for the other two Spanish Basque provinces of Alava and Vizcaya (Pérez-Miranda et al. 2005a, b) have been incorporated for comparison purposes. Several features associated with STR loci make them useful sites for the elucidation of human population history (Jorde et al. 1997; Shriver et al. 1997) and for studying genetic microdifferentiation among local subdivided populations (Reddy et al. 2001). These properties are large number of alleles, high H_o and abundance in the human genome, as well as technical considerations such as ease in genotyping and scoring (Zhivotovsky et al. 2004).

The most notable finding of this study is that the phylogenetic relationships resulting from the F_{ST} genetic distance matrix were strongly defined along ethnohistorical and geographical lines of the populations included in the analyses. Thus, the results of the present study indicate a clear genetic differentiation of native Basques, which separate them from the remaining populations of Europe (including Spain), the Middle East and North Africa. In contrast, no prominent genetic characteristic was found for RBAS, which plotted with the bulk of European populations. These findings could be ascribed to major demographic changes linked to the industrialization process in the Basque region, which propitiated the confluence and mixture of different Iberian populations in the Basque territories since as early as the first half of the nineteenth century (Alfonso-Sánchez et al. 2001). Demographic changes promoted by industrialization occurred mainly in zones close to the industrial centers; in rural zones, these demographic effects were practically negligible. In relation to this issue, we should stress that the group of resident Basques was collected in Bilbao, the most important industrial city in the Spanish Basque Country.

Based mainly on HLA data, previous works have associated the origin of Basques with a hypothetical ancient Berber settlement in the north of the Iberian Peninsula (Arnaiz-Villena et al. 1999). Interestingly, the

Table 6 Admixture contribution proportions from Guipúzcoan and Spanish groups to the gene pools of Vizcaya, Navarre, Alava, and residents of the Basque Country

	μ_1 (Guipúzcoan)	μ_2 (Spaniard)
Biscay	0.456 ± 0.055	0.544 ± 0.055
North Navarre	0.420 ± 0.049	0.580 ± 0.049
Alava	0.348 ± 0.060	0.652 ± 0.060
Residents	0.256 ± 0.030	0.744 ± 0.031

STR markers used in the present study reveal a remarkable genetic dissimilarity between autochthonous Basque groups (GUIP, NNAV, ALAV, and VIZC) and North African populations (EGYP, MORC). The asymmetrical partitioning of the STR diversity between Basques and North Africans is not supportive of a direct common ancestry and/or significant gene flow between these two regions. Therefore, the findings of the present study are not indicative of a paleo-North African origin for Basques; rather, our data provide new evidence on the low genetic affinity between both population groups, corroborating the conclusions of previous works (Bosch et al. 1997; Pérez-Miranda et al. 2003).

It is also worth noticing the remarkable genetic differentiation between native Basque groups and other European populations, especially those sharing the Iberian Peninsula. Usually, geographically close populations are also genetically close because of a common origin or extensive gene flow between them (Barbujani et al. 1994). However, Basques represent a group that is linguistically and genetically isolated within the Iberian Peninsula. The most common argument to account for the Basque distinctiveness is random genetic drift and inbreeding over long periods while isolated from surrounding populations. Yet, the causative agents of such marked isolation remain unclear. Some authors have suggested that the isolation of Basques is a consequence of their singular language (Cavalli-Sforza et al. 1994; Calderón et al. 1998; Pancorbo et al. 2001). Indeed, linguistic differences can be effective barriers to gene flow (Barbujani and Sokal 1990; Barbujani 1997).

A related issue is why the Basque language is restricted to the current Basque territory? Two major historical episodes might have played an important role in the shaping of the Iberian linguistic and/or genetic map: the Roman (BC 348–411 AD) and the Muslim (711–1492 AD) occupations of the Iberian Peninsula. It is well known, from historical evidence, that both Romanization and Arabization processes had only a minor impact in the Basque historical territories of the current provinces of Guipúzcoa and Vizcaya and the northernmost regions of the Alava and Navarre provinces (García de Cortázar 2004). The reasons for the lack of penetration into the Basque territory are not completely understood. Some findings from archaeological and paleoeconomic studies have associated the limited interest in Basque lands to the long-standing mainstay of the rest of the Iberian Peninsula based on extensive cultivation of cereals, grapevines, and olive trees (Apellániz 1975; Clark 1986).

The segregation and distance of the four Basque groups (GUIP, NNAV, VIZC, and ALAV) in both the NJ tree and in the MDS (see Figs. 2, 3) strongly suggest a lack of genetic homogeneity among the autochthonous Basque collections. This assumption seems to be confirmed by the AMOVA data. Evidence of the genetic heterogeneity among the Basques has been previously observed in studies on the variability of the immunoglobulin (GM and KM) genes (Calderón et al. 1998), on

the genetic polymorphism of HLA-DQA1 loci in different Basque samples (Pérez-Miranda et al. 2003), and on PAIs (Pancorbo et al. 2001). These studies suggest that the more probable cause of the genetic diversity among the Basque groups may be the existence of different levels of admixture of ancient Basques with other non-Basque neighboring populations. This argument is corroborated by the admixture proportions reported in this study. Bearing in mind the extremely reduced geographical distances between the traditional Basque territories, the genetic uniqueness and differences in the degree of isolation among the distinct autochthonous Basque groups are thought to be mainly conditioned by sociocultural features and in some areas by physical barriers in the form of deep, narrow valleys separated by mountain ranges.

Based on the diversity of the STR markers examined in the present study, Guipúzcoa exhibits the most genetic uniqueness of all four native Basque groups. Likewise, Basques from Alava drifted apart from the Basque cluster (Fig. 3). The provinces of Alava and Guipúzcoa have been considered as the two extremes of Basque genetic variation on the basis of classical polymorphisms (Manzano et al. 1996; Calderón et al. 1998). The STR data derived from NJ, MDS, and AMOVA analyses also indicate maximum genetic dissimilarity between the Alava and Guipúzcoa groups whereas Basques from Guipúzcoa and Northern Navarre show the greatest genetic affinity. The genetic similarity between the native populations of Guipúzcoa and North Navarre has been suggested in several studies (Calderón et al. 1998; Peña et al. 2002; Pérez-Miranda et al. 2003). The sample of Vizcaya tends to hold an intermediate position. All the above described coincide, for the most part, with results derived from admixture estimations. These findings indicate that the genetic partitioning inferred from the spatial variability of the STR diversity mirrors the current geographic distribution of the Basque language (*Euskera*). A recent report by the Basque government (1995) titled “The Continuity of the Basque Language” indicates that 44% of the present population of Guipúzcoa use *Euskera* as their usual form of communication (monolingual individuals) or use it occasionally (actively bilingual). In the northernmost part of Navarre, the percentage of Basque speakers reaches 40%. The corresponding figures in the rest of the Spanish Basque provinces are 24% in Vizcaya and 15% in Alava.

Prevalence of *Euskera* has no doubt contributed to the Basques’ relative genetic isolation. In rural areas of Guipúzcoa, Northern Navarre, and oriental regions of Vizcaya, the persistence of local populations strongly embedded in the traditional sociocultural mores of the autochthonous Basque society has allowed the maintenance of the Basque language. Such populations are predominantly concentrated in small villages (<2,000 inhabitants) where a deeply rooted farming economy still prevails close to industrial centers. Likewise, the Basques represent a special case of European population where consanguinity, closely related to sociocultural

characteristics, has traditionally been an important component of the marital structure (Alfonso-Sánchez et al. 2001, 2005).

Language can be a major sociocultural factor limiting gene flow and population admixture by preventing the integration of immigrants into the autochthonous population and by increasing ethnic endogamy, the main consequence of which would be the departure from panmixia (Alfonso-Sánchez et al. 2001). This effect may be direct or associated with other sociocultural differences, which in turn influence mating and/or dispersal of individuals. The effect of linguistic and geographic barriers would be to slow down progress toward genetic equilibrium by causing anisotropies in the mating and/or dispersal patterns (Barbujani and Sokal 1990; Barbujani 1997). This seems to be the case in rural Vizcaya, Northern Navarre, and especially Guipúzcoa where the use of the Basque language and other shared sociocultural characteristics could have acted as barriers to random mating. Some recent findings appear to indicate that within the autochthonous groups of some Basque territories, there is a great reluctance to truncate the social and cultural patterns that promote close consanguinity (Alfonso-Sánchez et al. 2001). The results presented herein corroborate this hypothesis.

In short, the Basques' STR diversity revealed a substantial geographical partitioning, which seems to be the consequence of the following major factors: (1) language boundaries due to linguistic differences within the Basque area resulting from the differential impact of Latin and derived Romance languages (Michelena 1964), (2) more recently, socioeconomic and demographic aspects related to differences in the chronology and intensity of industrialization (different demographic structures among regions caused mainly by long-standing, localized immigration), and (3) the combined effects of the cited factors in the characteristic marital structure of each Basque territory (see Alfonso-Sánchez et al. 2001, 2005). The interaction among these variables may have led to the spatially structured genetic heterogeneity found in the contemporary autochthonous Basque population. In addition, the results of the present study underscore the usefulness and reliability of STRs for personal identity testing as expressed in the markedly high values obtained for both PD and PIC.

Acknowledgements A.M. Pérez-Miranda was supported by a postdoctoral fellowship MECD/Fulbright (Ministerio de Educación, Cultura y Deporte, Spain). M.A. Alfonso-Sánchez was supported through a postdoctoral fellowship of the Programa de Formación de Investigadores, Departamento de Educación, Universidades e Investigación (Basque government).

References

- Abdin L, Shimada I, Brinkmann B, Hohoff C (2003) Analysis of 15 short tandem repeats reveals significant differences between the Arabian populations from Morocco and Syria. *Legal Med* 5:S150–S155
- Aguirre AI, Vicario A, Mazón LI, Estomba A, de Pancorbo MM, Arrieta-Pico V, Pérez-Elortondo F, Lostao CM (1991) Are the Basques a single and unique population? *Am J Hum Genet* 49:450–458
- Akbasak BS, Budowle B, Reeder DJ, Redman J, Kline MC (2001) Turkish population data with the CODIS multiplex short tandem repeat loci. *Forensic Sci Int* 123:227–229
- Alfonso-Sánchez MA, Peña JA, Aresti U, Calderón R (2001) An insight into recent consanguinity within the Basque area in Spain. Effects of autochthony, industrialization and demographic changes. *Ann Hum Biol* 28:505–521
- Alfonso-Sánchez MA, Aresti U, Peña JA, Calderón R (2005) Inbreeding levels and consanguinity structure in the Basque province of Guipúzcoa (1862–1980). *Am J Phys Anthropol* 127:240–252
- Alonso S, Armour JA (1998) MS 205 minisatellite diversity in Basques: evidence for a pre-Neolithic component. *Genome Res* 8:1289–1298
- Alshamali FH, Alkhayat AI, Budowle B, Watson ND (2003) Allele frequency distributions and other population genetic parameters for 13 STR loci in a UAE local population from Dubai. *Int Congr Ser* 1239:249–258
- Apellániz JM (1975) El grupo de Santimamiñe durante la Prehistoria con cerámica. *Munibe* 27:1–136
- Arnaiz-Villena A, Martínez-Laso J, Alonso-García S (1999) Iberia: population genetics, anthropology, and linguistics. *Hum Biol* 71:725–743
- Barbujani G (1997) DNA variation and language affinities. *Am J Hum Genet* 61:1011–1014
- Barbujani G, Sokal RR (1990) Zones of abrupt genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA* 87:1816–1819
- Barbujani G, Nasidze IS, Whitehead GN (1994) Genetic diversity in the Caucasus. *Hum Biol* 66:639–668
- Basque Government (1995) La continuidad del Euskera. Servicio Central de Publicaciones del Gobierno Vasco, Vitoria
- 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* 59:63–81
- Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, Bertranpetit J (1997) Population history of North Africa. Evidence from classical genetic markers. *Hum Biol* 69:295–311
- Brown RJ, Rowold D, Tahir M, Barna C, Duncan G, Herrera RJ (2000) Distribution of the HLA-DQA1 and polymarker alleles in the Basque population of Spain. *Forensic Sci Int* 108:145–151
- Calafell F, Bertranpetit J (1994) Principal component analysis of gene frequencies and the origin of Basques. *Am J Phys Anthropol* 93:201–215
- Calderón R, Vidales C, Peña JA, Pérez-Miranda AM, Dugoujon JM (1998) Immunoglobulin allotypes (GM and KM) in Basques from Spain: approach to the origin of the Basque population. *Hum Biol* 70:667–698
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, New Jersey
- Clark GA (1986) El nicho alimentario humano en el norte de España desde el Paleolítico hasta la romanización. *Trabajos de Prehistoria* 43:159–184
- Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bertranpetit J (1998) HLA evidence for the lack of genetic heterogeneity in Basques. *Ann Hum Genet* 62:123–132
- DNA recommendations (1994) Report concerning further recommendations of the DNA Commission of the ISFH regarding PCR based polymorphisms in STR (short tandem repeat) systems. *Int J Legal Med* 107:159–160
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791

- Felsenstein J (1989) PHYLIP: Phylogeny inference package (Version 3.2). *Cladistics* 5:164–166
- García O, Uriarte I, Peñas R, Martín P, Albarrán C, Alonso A (2003) The CODIS system in the Basque Country resident population studied with multiplex systems. *Int Congr Ser* 1239:193–196
- García de Cortázar F (2004) *Memoria de España*. Editorial Aguilar, Madrid
- García-Hirschfeld J, Farfan MJ, Prieto V, López-Soto M, Torres Y, Sanz P (2003) Allelic distribution of 15 STRs in a population from Extremadura (Central-Western Spain). *Int Congr Ser* 1239:165–169
- Garofano L, Pizzamiglio M, Vecchio C, Lago G, Floris T, D'Er-rico G, Brembilla G, Romano A, Budowle B (1998) Italian population data on thirteen short tandem repeat loci: HUM-TH01, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358. *Forensic Sci Int* 97:53–60
- Goedde HW, Hirth L, Benkmann HG, Pellicer A, Pellicer T, Stahn M, Singh S (1972) Population genetic studies of red cell enzyme polymorphisms in four Spanish populations. *Hum Hered* 22:552–560
- Goedde HW, Hirth L, Benkmann HG, Pellicer A, Pellicer T, Stahn M, Singh S (1973) Population genetic studies of serum protein polymorphisms in four Spanish populations. *Hum Hered* 23:135–146
- Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48:361–372
- Hearne C, Ghosh S, Todd J (1992) Microsatellites for linkage analysis of genetic traits. *Trend Genet* 8:288–294
- Iriondo M, Barbero MC, Manzano C (2003) DNA polymorphisms detect ancient barriers to gene flow in Basques. *Am J Phys Anthropol* 122:73–84
- Jorde LB, Rogers AR, Bamshad M, Scott Watkins W, Krakowiak P, Sung S, Kere J, Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA* 94:3100–3103
- Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27
- Long JC, Williams RC, McAuley JE, Medis R, Partel R, Tregellas WM, South SF, Rea AE, McCormick SB, Iwaniec U (1991) Genetic variation in Arizona Mexican Americans: Estimation and interpretation of admixture proportions. *Am J Phys Anthropol* 84:141–157
- Lucotte G, Hazout S (1996) Y-chromosome DNA haplotypes in Basques. *J Mol Evol* 42:472–475
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning. A laboratory manual*. Cold Spring Harbor Laboratory Publications, New York, pp 458–459
- Manzano C, Aguirre AI, Iriondo M, Martín M, Osaba L, de la Rúa C (1996) Genetic polymorphisms of the Basques from Gipuzkoa: genetic heterogeneity of the Basque population. *Ann Hum Biol* 23:285–296
- Manzano C, de la Rúa C, Iriondo M, Mazón LI, Vicario A, Aguirre AI (2002) Structuring the genetic heterogeneity of the Basque population: a view from classical polymorphisms. *Hum Biol* 74:51–74
- Michelena L (1964) *Sobre el pasado de la lengua vasca*. Ediciones Auñamendi, San Sebastián
- Mourant AE (1947) The blood groups of the Basques. *Nature* 160:505
- de Pancorbo MM, López-Martínez M, Martínez-Bouzas C, Castro A, Fernández-Fernández I, Antúnez de Mayolo G, Antúnez de Mayolo A, Antúnez de Mayolo P, Rowold DJ, Herrera RJ (2001) The Basques according to polymorphic Alu insertions. *Hum Genet* 109:224–233
- Paredes M, Crespillo M, Luque JA, Valverde JL (2003) STR frequencies for the PowerPlex 16 System Kit in a population from Northeast Spain. *Forensic Sci Int* 135:75–78
- Pawlowski R, Maciejewska A (2000) The forensic validation studies of Profiler Plus and allele frequencies of profiler loci in a polish population. *Prog Forensic Genet* 8:136–138
- Peña JA, Calderón R, Pérez-Miranda AM, Vidales C, Dugoujon JM, Carrión M, Crouau-Roy B (2002) Microsatellite DNA markers from HLA region (D6S105, D6S265 and TNFa) in autochthonous Basques from Northern Navarre (Spain). *Ann Hum Biol* 29:176–191
- Pérez-Miranda AM, Alfonso-Sánchez MA, Peña JA, Calderón R (2003) HLA-DQA1 polymorphism in autochthonous Basques from Navarre (Spain): genetic position within European and Mediterranean scopes. *Tissue Antigens* 61:465–474
- Pérez-Miranda AM, Alfonso-Sánchez MA, Peña JA, Pancorbo MM de, Herrera RJ (2005a) Genetic polymorphisms at 13 STR loci in autochthonous Basques from Alava province (Spain). *Legal Med* 7:58–61
- Pérez-Miranda AM, Alfonso-Sánchez MA, Kalantar A, Peña JA, Pancorbo MM de, Herrera RJ (2005b) Allele frequencies of 13 STR loci in autochthonous Basques from the province of Vizcaya (Spain). *Forensic Sci Int* 152:259–262
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Reddy BM, Sun G, Luis JR, Crawford MH, Hemam NS, Deka R (2001) Genomic diversity at thirteen short tandem repeat loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. *Hum Biol* 73:175–190
- Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767–779
- Rowold DJ, Herrera RJ (2003) Inferring recent phylogenies using forensic STR technology. *Forensic Sci Int* 133:260–265
- Rowold DJ, Herrera RJ (2005) On human STR sub-population structure. *Forensic Sci Int* 151:59–69
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sanz P, Prieto V, Flores I, Torres Y, López-Soto M, Farfan MJ (2001) Population data of 13 STRs in southern Spain (Andalusia). *Forensic Sci Int* 119:113–115
- Schneider S, Roessli D, Excoffier L (2000) *A software for population genetics data analysis. Arlequin Version 2.000*. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Shriver MD, Jin L, Ferrell RE, Deka R (1997) Microsatellite data support an early population expansion in Africa. *Genome Res* 7:586–591
- Smouse PE, Chakraborty R (1986) The use of restriction fragment length polymorphisms in paternity analysis. *Am J Hum Genet* 38:918–939
- Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 74:50–61