Autosomal STR Variation in a Basque Population: Vizcaya Province

M. ZLOJUTRO,[1] R. ROY,[2] J. PALIKIJ,[1] AND M. H. CRAWFORD[1]

Abstract We have characterized 68 unrelated Basque individuals from Vizcaya, Spain, for 13 tetrameric short tandem repeat (STR) loci: *CSF1PO*, *D3S1358*, *D5S818*, *D7S820*, *D8S1179*, *D13S317*, *D16S539*, *D18S51*, *D21S11*, *FGA*, *TH01*, *TPOX*, and *VWA*. Interpopulational analyses were also performed for 21 European and North African population data sets for nine of the STRs typed in the Basque sample. Heterozygosity values for the Vizcayan Basques were found to be high, ranging from 0.662 to 0.882, and none of the STR loci significantly deviated from Hardy-Weinberg equilibrium. Based on the comparative population data set, the average G_{ST} score is 0.7%, indicating a low degree of genetic differentiation. However, neighbor-joining trees and multidimensional-scaling plots of D_A genetic distances indicate that the Vizcayan Basques are an outlier relative to both neighboring Iberians and North African populations.

The Basques are a European population residing in the western portion of the Pyrenees. They are considered by many anthropologists and linguists to represent a relic or "living fossil" of ancestral Paleolithic Europeans (Cavalli-Sforza et al. 1994). This long-held perception partly stems from the distinctiveness of the Basque language, Euskera, which is a non-Indo-European isolate within the relatively homogeneous Indo-European landscape of western Europe and whose linguistic classification is yet to be resolved with any degree of certainty (Trask 1997). From the genetic perspective the Basques were first characterized for the ABO blood system (Boyd and Boyd 1937), revealing an unusual pattern of a high frequency of type O blood and the lowest frequency of type B in Europe.

[1]Department of Anthropology, University of Kansas, Lawrence, KS 66045.
[2]St. Louis County Police Crime Laboratory, St. Louis, MO 63105.

We are grateful to all the Basque donors for participating in the present study and to Arantza Apraiz for the collection of the buccal samples. We thank Phil Melton, Alan Redd, and two anonymous reviewers for their critical reading of the manuscript and helpful critiques. This research was supported by the National Geographic Society through grant 6935-00.

Human Biology, October 2006, v. 78, no. 5, pp. 599–618.

KEY WORDS: SHORT TANDEM REPEATS (STRs), *CSF1PO*, *D3S1358*, *D5S818*, *D7S820*, *D8S1179*, *D13S317*, *D16S539*, *D18S51*, *D21S11*, *FGA*, *HUMTH01*, *TPOX*, *VWA*, BASQUE, VIZCAYA, GENETIC STRUCTURE, MEDITERRANEAN BASIN POPULATIONS.

Subsequent studies of other classical polymorphisms also found atypical frequencies of various alleles, which established the Basques as a genetic outlier within the European gene pool (Etcheverry 1945; Chalmers et al. 1949; Vergnes et al. 1980; Davrinche et al. 1981; Aguirre et al. 1991; Bertranpetit and Cavalli-Sforza 1991; Manzano et al. 1993; Calafell and Bertranpetit 1994; Esparza et al. 1995; Iriondo et al. 1996; Calderón et al. 1998).

However, recent molecular research involving mitochondrial and Y-chromosome DNA has not supported the genetic uniqueness of the Basques. mtDNA variability among Europeans displays low population substructuring (Richards et al. 1996, 2000), with the Basques sharing common haplogroups with other Iberian populations that date to both the Neolithic and the Paleolithic periods. Torroni et al. (1998) reported a high frequency (20%) of haplogroup V in the Basques, which contrasts with its lower frequency in other European populations, and argued that these results reflect a population expansion after the Last Glacial Maximum from reduced gene pools located in southwestern France and Cantabria. But the findings of other mtDNA studies are not consistent with this model; they report smaller frequencies of haplogroup V in the Basques [e.g., 3% by Côrte-Real et al. (1996)] and moreover fail to identify this particular lineage among prehistoric Basque specimens (Izagirre and de la Rúa 1999). Similarly, for Y-chromosome markers, the Basques share with the rest of Europeans the most common haplogroups and haplotypes and are not significantly differentiated from neighboring populations (Alonso et al. 2005).

Basque genetic variability has also been characterized for autosomal microsatellites or short tandem repeats (STRs). Since their first description in the early 1990s (Edwards et al. 1991; Polymeropoulos et al. 1991), STRs have been actively used in elucidating the evolutionary history of human populations (Bowcock et al. 1994; Jorde et al. 1997). STRs are short sequences of DNA composed of repeated motifs consisting of two to six nucleotides that are ubiquitously distributed in the genome. Commercial multiplex PCR kits and capillary sequencers have allowed for rapid genotyping of multiple autosomal STR loci, producing highly informative population data sets that are less influenced by stochastic evolutionary processes at any single locus. Moreover, genetic data based on autosomes are likely to reflect the demographic history and migratory patterns of both sexes and thus produce an intermediate picture relative to data based solely on mtDNA or Y-chromosome polymorphisms. Several studies characterizing autosomal STR diversity in the Basques have been conducted to date, each varying with regard to the loci analyzed and the geographic location of the population sample (Pérez-Lezaun et al. 1997; Bosch et al. 2000; Garcia et al. 2001; Peña et al. 2002; Iriondo et al. 2003). Consistent with earlier population genetic studies, the Basques tend to differ from neighboring European populations based on STR data.

In this paper we contribute to the growing body of genetic information on the Basque people by characterizing allele-frequency distributions at 13 STR loci for a Basque sample collected from the Spanish province of Vizcaya, and we

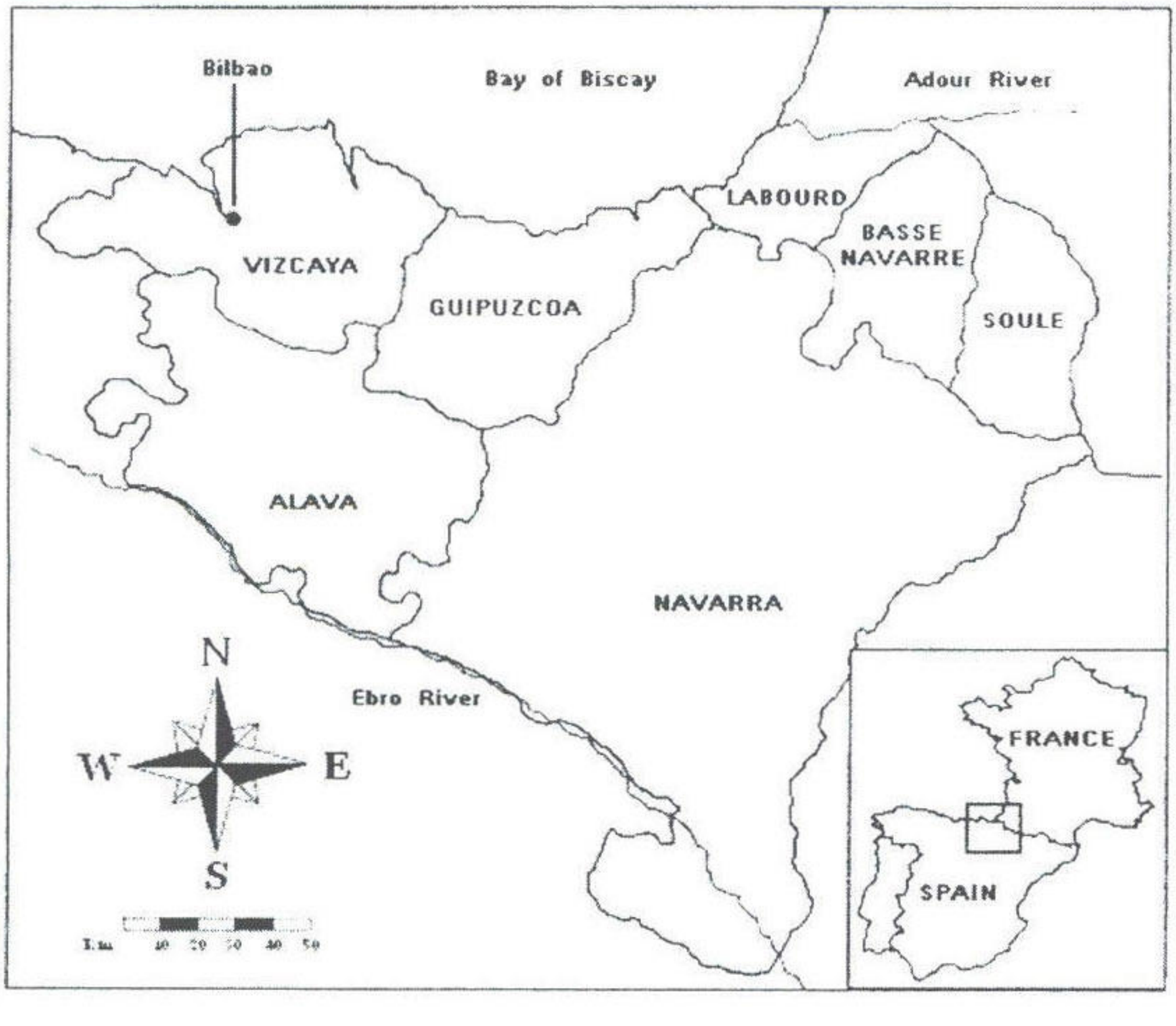

Figure 1. Geographic locations of Bilbao and the provinces representing the Basque areas of Spain and France.

perform interpopulational analyses with other European and North African groups from the Mediterranean basin for nine of the selected STR loci to evaluate the genetic distinctiveness of the Basques, if any.

Materials and Methods

Sampling. Buccal swabs were collected from 68 unrelated Basque individuals from the Basauri suburb of Bilbao, Spain. Appropriate informed consent was obtained, and the participants were self-identified as being of Basque ancestry. Figure 1 shows the Basque-speaking provinces of Spain and France and the location of Bilbao, the capital of Vizcaya. The buccal samples were stored at −20°C. DNA was extracted from the samples using a standard phenol-chloroform technique (Birren et al. 1997).

DNA Analysis. At the St. Louis County Crime Laboratory (St. Louis, Missouri), the DNA samples were analyzed for 13 STR loci: *CSF1PO*, *D3S1358*,

Table 1. STR Information

STR Locus	*Chromosome Location*	*Repeat Sequence*[a]	*Reference*
CSF1PO	5q33.3–34	AGAT	Hammond et al. (1994)
D3S1358	3p	AGAT	Li et al. (1993)
D5S818	5q21–31	AGAT	Hudson et al. (1995)
D7S820	7q	GATA	Green et al. (1991)
D8S1179	8	TATC	Oldroyd et al. (1995)
D13S317	13q22–31	TATC	Hudson et al. (1995)
D16S539	16q22–24	GATA	Cooperative Human Linkage Center[b]
D18S51	18q21.3	GAAA	Urquhart et al. (1995)
D21S11	21	TCTA	Sharma and Litt (1992)
FGA	4q28	CTTT	Mills et al. (1992)
TH01	11p15–15.5	TCAT	Edwards et al. (1992)
TPOX	2p23–pter	AATG	Anker et al. (1992)
VWA31A	12p12–pter	TCTA	Kimpton et al. (1992)

a. Top strand (5′ → 3′).
b. Genebank accession number G07925.

D5S818, *D7S820*, *D8S1179*, *D13S317*, *D16S539*, *D18S51*, *D21S11*, *FGA*, *TH01*, *TPOX*, and *VWA31A*. Table 1 provides the chromosome location of each STR locus, the repeat sequence motif, and the corresponding reference. The selected loci are situated on 12 different chromosomes, with both *CSF1PO* and *D5S818* found on chromosome 5. Multiplex amplifications were accomplished using the AmpFlSTR Profiler Plus and COfiler kits (Applied Biosystems, Foster City, California). PCR reactions were prepared according to the manufacturer's recommended protocol and performed on a Perkin-Elmer GeneAmp PCR System 9700 thermal cycler. PCR products were separated by capillary electrophoresis using the ABI Prism 310 Genetic Analyzer (Applied Biosystems). Output data were compiled and analyzed using GeneScan and Genotyper accessory software, resulting in export files containing allele tables for each STR marker.

Comparative Population Data. Allele frequency data for nine of the STR loci typed in the present study were obtained from the literature for 20 European and North African populations (Table 2). This comparative data set includes an additional Basque sample that was collected in several towns and villages within the Basque province of Guipuzcoa (Pérez-Lezaun et al. 2000).

Statistical Analysis. Allele frequencies for the Vizcayan sample were determined by a simple gene-counting method (Li 1976). Using Arlequin, version 2.0 (Schneider et al. 2000), we calculated observed and expected heterozygosities for the 13 STR markers, and we evaluated deviations from Hardy-Weinberg equilibrium using the exact test algorithm described by Guo and Thompson (1992).

The coefficient of gene differentiation, $G_{ST} = (H_T - H_S)/H_T$, was computed for each locus in the comparative population data set using the computer

Table 2. Populations Included in Interpopulational Analyses Based on Nine STR Loci

Population	*Abbreviation*	*Sample Size (2n)*	*Reference*
Vizcaya	VIZC	136	Present study
Guipuzcoa	GUPZ	200	Pérez-Lezaun et al. (2000)
Cantabria	CNTB	316	Zarrabeitia and Riancho (2001)
Catalonia	CATL	200	Pérez-Lezaun et al. (2000)
Galicia	GALC	244	Gusmão et al. (2000)
Valencia	VALN	202	Tomás et al. (2001)
Murcia	MURC	228	Toria et al. (2002)
Andalusia	ANDL	394	Sanz et al. (2001)
Portugal	PORT	854	Amorim et al. (2001)
Balearic Islands	BLRC	226	Tomás et al. (2000)
Madeira	MADR	200	Fernandes et al. (2002)
Canary Islands	CNRY	276	Gamero et al. (2000)
Morocco	MORC	372	Jauffrit et al. (2003)
Maghreb	MGHB	236	Farfán et al. (2001)
Tuscany	TSCY	376	Ricci et al. (2002)
Greece	GREC	286	Sánchez-Diz et al. (2002)
Turkey	TURK	620	Asicioglu et al. (2002)
Egypt	EGYP	282	Klintschar et al. (1999)
Germany	GERM	310	Anslinger et al. (2001)
Austria	AUST	388	Neuhuber et al. (1999)
Poland	PLND	826	Pepinski et al. (2001)

program DISPAN (Ota 1993), where H_T is the gene diversity for the total population (i.e., average allele frequencies for the entire data set) and H_S is the average of the gene diversities computed for the individual (sub-) populations (Nei 1987).

D_A genetic distances (Nei et al. 1983) were computed between populations using the Phylip software package, version 3.6 (Felsenstein 1993). Relative to other distance measures, D_A is least affected by sample size and was found to produce the most accurate phylogenetic trees under various evolutionary conditions (Takezaki and Nei 1996). A neighbor-joining tree was constructed using D_A distances (Saitou and Nei 1987), and its robustness was established using bootstrap resampling procedures. Because a tree representation of genetic distances can be misinterpreted as a succession of population splits, the D_A matrix was also projected in two-dimensional space by multidimensional scaling using the NTSYS statistical software (Rohlf 2002).

To assess the gene flow experienced by these populations, the r_{ii} score (i.e., the genetic distance of a population from the centroid) was estimated by calculating the mean value for the following expression:

$$r_{ii} = \frac{(p_i - \overline{p})^2}{\overline{p}(1-\overline{p})}, \tag{1}$$

where r_{ii} is the distance from the centroid for the ith population for a particular allele, p_i is the frequency of the allele within the ith population, and $\overline{p}$ is the mean

allele frequency for the entire population data set. A regression model of the mean population heterozygosities against the corresponding r_{ii} scores was produced using the program Minitab, version 12.0 (Minitab Inc., State College, Pennsylvania). According to Harpending and Ward (1982), a linear relationship is expected between heterozygosity and genetic distance, and deviations from this simple relationship can be interpreted as evidence of either increased gene flow or greater population isolation, depending on where a population lies with respect to the theoretical regression line.

Results

Allele Frequency and Genomic Diversity Within the Basque Population. The allele frequency distributions for the 13 STR loci are shown in Table 3. The majority follow a unimodal pattern, although six systems do exhibit varying degrees of multimodality. Observed heterozygosity scores range from 0.662 to 0.882, with the extreme values represented by *TH01* and *D18S51*, respectively. None of the STR loci display significant deviation from Hardy-Weinberg equilibrium for the Vizcayan sample, suggesting random mating and absence of genetic substructure involving these loci.

Genomic Diversity Between Populations. To measure the degree of genetic differentiation among the Basques and the selected comparative populations, we calculated G_{ST} values (a measure of interpopulational variability) for nine STR loci. The gene diversity results for each locus are presented in Table 4. The total genomic diversity (H_T) among the populations is high, ranging from 0.726 for *D5S818* to 0.877 for *D18S51*. However, most of the genomic diversity is accounted by the variability between individuals within populations (H_S). The percentage of genomic diversity attributable to variability between populations relative to the total genomic diversity (G_{ST}) ranges from a high of 1.0% for *D13S317* to a low of 0.5% for *VWA*. For the entire data set, 0.7% of the total genomic diversity is attributable to variability between populations.

Genomic Affinities Among Populations. The smallest pairwise D_A genetic distances between the Vizcayan Basques and other reference populations from the Mediterranean basin are with Murcia from southeastern Spain (0.0187), the neighboring province of Cantabria (0.0203), and Portugal (0.0204). The largest genetic distance is with the sample from Catalonia (0.0332). For the Basque sample from Guipuzcoa the smallest pairwise genetic distances are with the Vizcayan Basques (0.0300) and the Cantabrians (0.0326). The D_A matrix (not shown) was used to construct a neighbor-joining tree (Figure 2) with bootstrap values at the nodes representing 1,000 randomizations of the population frequency data. The Vizcayans cluster with the Guipuzcoans at a robust 70%, although they are separated by long phylogenetic branches that reflect the large D_A score for the two Basque samples. The Cantabrians cluster with the Basques in

Table 3. Allele Frequency Distributions at 13 STR Loci Analyzed for the Vizcayan Basque Sample

Allele	*FGA*	*VWA*	*D3S1358*	*D5S818*	*D7S820*	*D8S1179*	*D13S317*	*D18S51*	*D21S11*	*CSF*	*TH01*	*TPOX*	*D16S539*
6											0.300		
7					0.008						0.046		
8					0.103	0.037	0.156				0.115	0.425	
9				0.029	0.159	0.037	0.031			0.008	0.177	0.127	0.117
9.3											0.354		
10				0.132	0.373	0.074	0.117	0.037		0.367	0.008	0.067	0.047
11				0.353	0.222	0.066	0.281	0.022		0.227		0.351	0.266
12				0.390	0.103	0.096	0.273	0.202		0.313		0.030	0.352
13				0.096	0.024	0.302	0.109	0.112		0.063			0.195
14		0.103	0.119		0.008	0.250	0.031	0.202		0.016			0.023
15		0.125	0.269			0.110		0.127		0.008			
16		0.235	0.224			0.029		0.127					
17		0.272	0.149					0.082					
18	0.030	0.154	0.231					0.022					
19	0.097	0.110	0.008					0.008					
20	0.172							0.045					
21	0.112							0.015					
22	0.112												
22.2	0.008												
23	0.239												

Table 3. (Continued)

Allele	*FGA*	*VWA*	*D3S1358*	*D5S818*	*D7S820*	*D8S1179*	*D13S317*	*D18S51*	*D21S11*	*CSF*	*TH01*	*TPOX*	*D16S539*
24	0.127												
25	0.090												
26	0.008												
27									0.030				
28	0.008								0.090				
29									0.202				
30									0.299				
30.2									0.045				
31									0.045				
31.2									0.082				
32.2									0.112				
33									0.008				
33.2									0.075				
34.2									0.015				
N	67	68	67	68	63	68	64	67	67	64	65	67	64
Observed heterozygosity	0.836	0.824	0.776	0.765	0.778	0.838	0.813	0.882	0.881	0.688	0.662	0.702	0.719
Expected heterozygosity	0.865	0.817	0.795	0.707	0.772	0.818	0.805	0.869	0.839	0.722	0.744	0.680	0.764
P^a	0.533	0.581	0.162	0.601	0.574	0.685	0.569	0.777	0.542	0.591	0.576	0.489	0.213

a. Guo and Thompson (1992) test for deviation from Hardy-Weinberg equilibrium.

Table 4. Gene Diversity Analysis of Nine STR Loci for European and North African Populations

Population	*FGA*	*VWA*	*D3S1358*	*D5S818*	*D7S820*	*D8S1179*	*D13S317*	*D18S51*	*D21S11*	*Average*
Vizcaya	0.860	0.814	0.793	0.701	0.770	0.818	0.800	0.869	0.839	0.807
Guipuzcoa	0.850	0.790	0.788	0.748	0.759	0.803	0.771	0.860	0.794	0.796
Cantabria	0.871	0.803	0.796	0.715	0.796	0.826	0.779	0.882	0.846	0.813
Catalonia	0.860	0.825	0.785	0.714	0.815	0.781	0.769	0.879	0.809	0.804
Galicia	0.855	0.822	0.786	0.712	0.796	0.817	0.795	0.880	0.830	0.810
Valencia	0.872	0.807	0.800	0.702	0.803	0.826	0.781	0.875	0.839	0.812
Murcia	0.860	0.820	0.815	0.718	0.787	0.807	0.758	0.866	0.826	0.806
Andalucía	0.868	0.805	0.803	0.708	0.797	0.824	0.795	0.879	0.857	0.815
Portugal	0.862	0.810	0.786	0.710	0.811	0.816	0.785	0.876	0.848	0.811
Balearic Islands	0.868	0.819	0.785	0.711	0.817	0.823	0.771	0.882	0.844	0.813
Madeira	0.863	0.813	0.793	0.746	0.810	0.810	0.770	0.868	0.838	0.812
Canary Islands	0.859	0.823	0.798	0.726	0.818	0.831	0.796	0.884	0.823	0.818
Morocco	0.851	0.822	0.779	0.727	0.772	0.824	0.748	0.878	0.831	0.803
Maghreb	0.842	0.793	0.772	0.734	0.775	0.833	0.720	0.877	0.846	0.799
Tuscany	0.865	0.793	0.788	0.722	0.796	0.832	0.745	0.868	0.852	0.807
Greece	0.855	0.822	0.788	0.733	0.794	0.814	0.775	0.881	0.846	0.812
Turkey	0.864	0.802	0.780	0.751	0.813	0.822	0.779	0.872	0.842	0.814
Egypt	0.873	0.806	0.772	0.763	0.784	0.819	0.792	0.858	0.825	0.810
Germany	0.870	0.818	0.781	0.709	0.814	0.790	0.774	0.885	0.841	0.809
Austria	0.864	0.806	0.806	0.709	0.808	0.814	0.801	0.872	0.854	0.815
Poland	0.863	0.804	0.802	0.717	0.812	0.797	0.759	0.873	0.866	0.810
H_S [a]	0.859	0.809	0.788	0.720	0.795	0.813	0.772	0.871	0.834	0.807
H_T [b]	0.864	0.813	0.793	0.726	0.800	0.820	0.780	0.877	0.841	0.813
G_{ST} [c]	0.007	0.005	0.007	0.008	0.007	0.008	0.010	0.007	0.008	0.007

a. Gene diversity within subpopulations.
b. Gene diversity among subpopulations.
c. Coefficient of gene differentiation (Nei 1973).

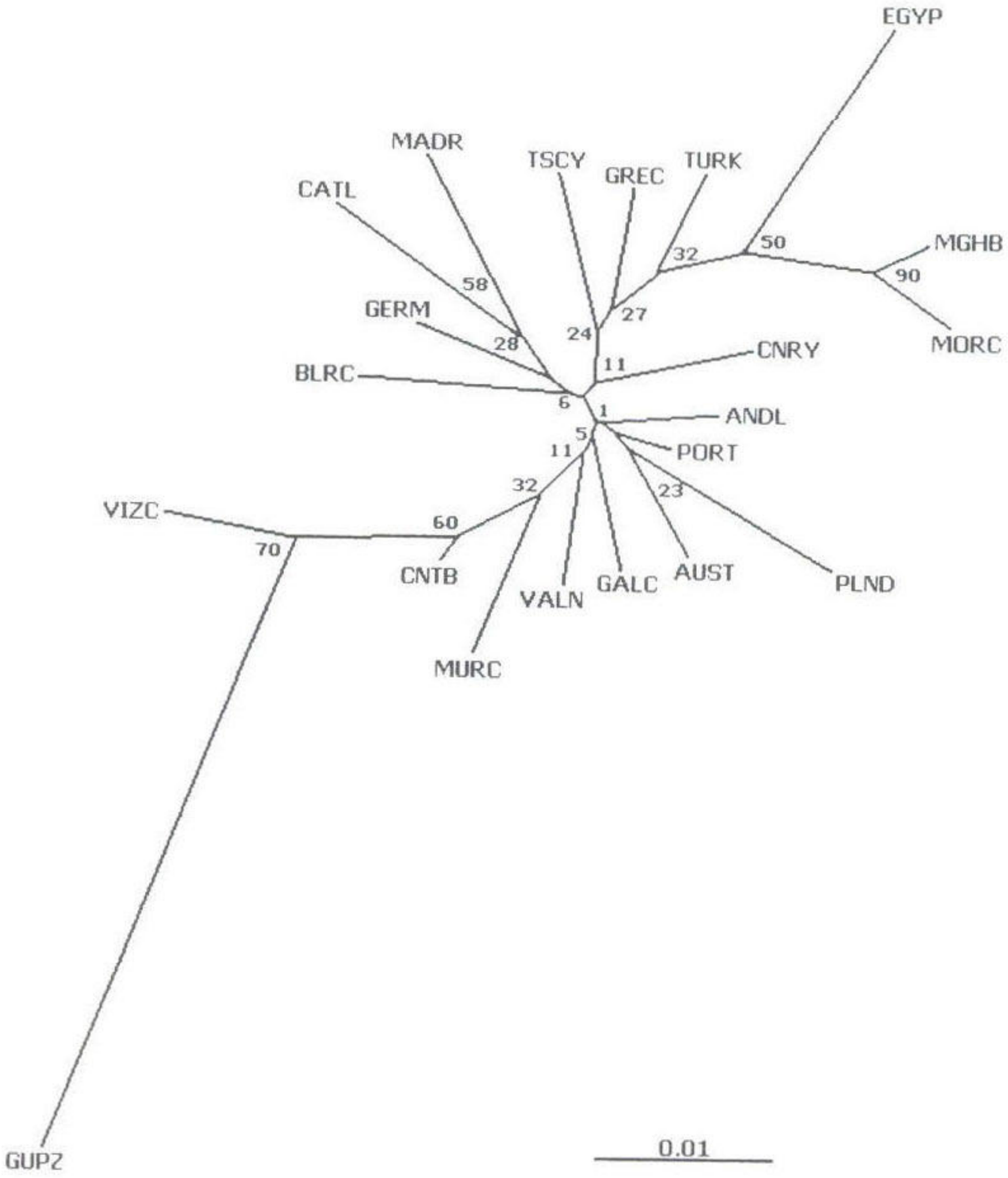

Figure 2. Unrooted neighbor-joining tree constructed from D_A genetic distances based on allele frequencies for nine STR loci in European and North African populations (1,000 bootstrap iterations). See Table 2 for population abbreviations.

60% of the bootstrapped trees. Situated at the opposing end of the neighbor-joining tree are the North Africans, who exhibit relatively high bootstrap scores, and populations from the eastern Mediterranean basin (Greeks and Anatolian Turks). The remaining Iberian populations cluster among themselves and with other European groups.

A multidimensional-scaling plot of the D_A distance matrix is presented in Figure 3. The stress value for the plot is 0.1851, which indicates a fair to poor goodness-of-fit to the original distance matrix (Kruskal 1964). Against the first dimension the Guipuzcoans are clearly differentiated from other comparative populations. The Vizcayans are also removed from the other populations relative

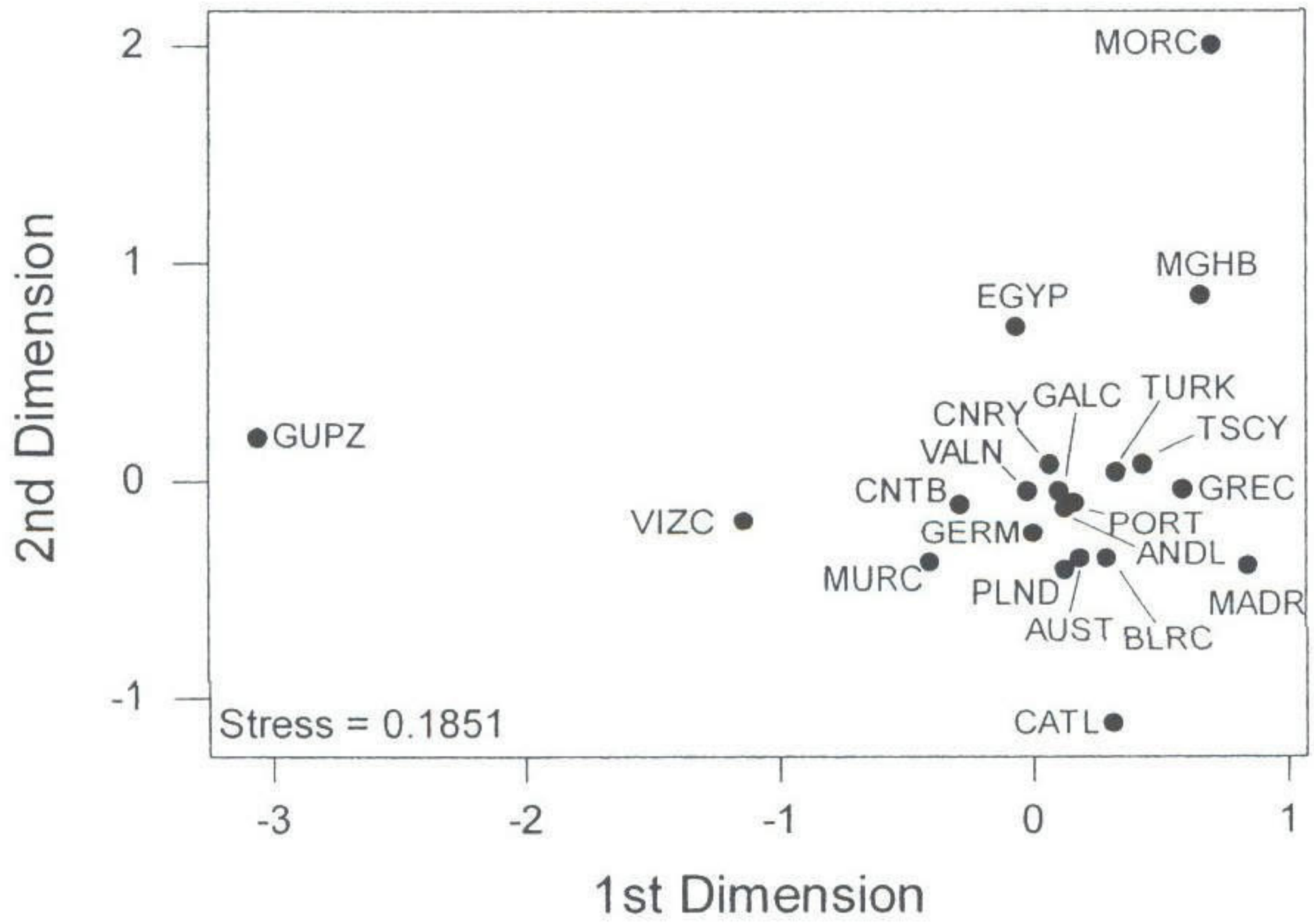

Figure 3. Multidimensional scaling applied to D_A genetic distances based on allele frequencies for nine STR loci in European and North African populations. See Table 2 for population abbreviations.

to the first dimension, but to a lesser degree. Along the second axis the North Africans are separated from the European groups. Overall, two population groupings are evident from the two-dimensional plot: a tight cluster of European populations and the more dispersed North Africans, with the Basque samples appearing as outliers.

To provide a more detailed picture of the genetic relationships among western Mediterranean peoples, the Iberian groups within the comparative population data set were analyzed separately. A neighbor-joining tree for this population subset is displayed in Figure 4. Once again, the two Basque samples couple together at a robust 88%, with the Guipuzcoans isolated by a long phylogenetic branch. The next closest population is the Cantabrians (70%), followed by Murcia (70%). The multidimensional-scaling plot for the circumscribed data set is presented in Figure 5. The Guipuzcoa sample was excluded from this analysis in order to focus on the relationship between the Vizcayan sample and the Iberian groups. The stress score is 0.0589, indicating a good fit to the distance matrix. Along the first dimension the Vizcayans are clearly differentiated from other Iberian populations. Against the second dimension the Andalusians and Murcians represent the extreme positions along this axis. Overall, the Vizcayans appear to be a genetic outlier among Iberians.

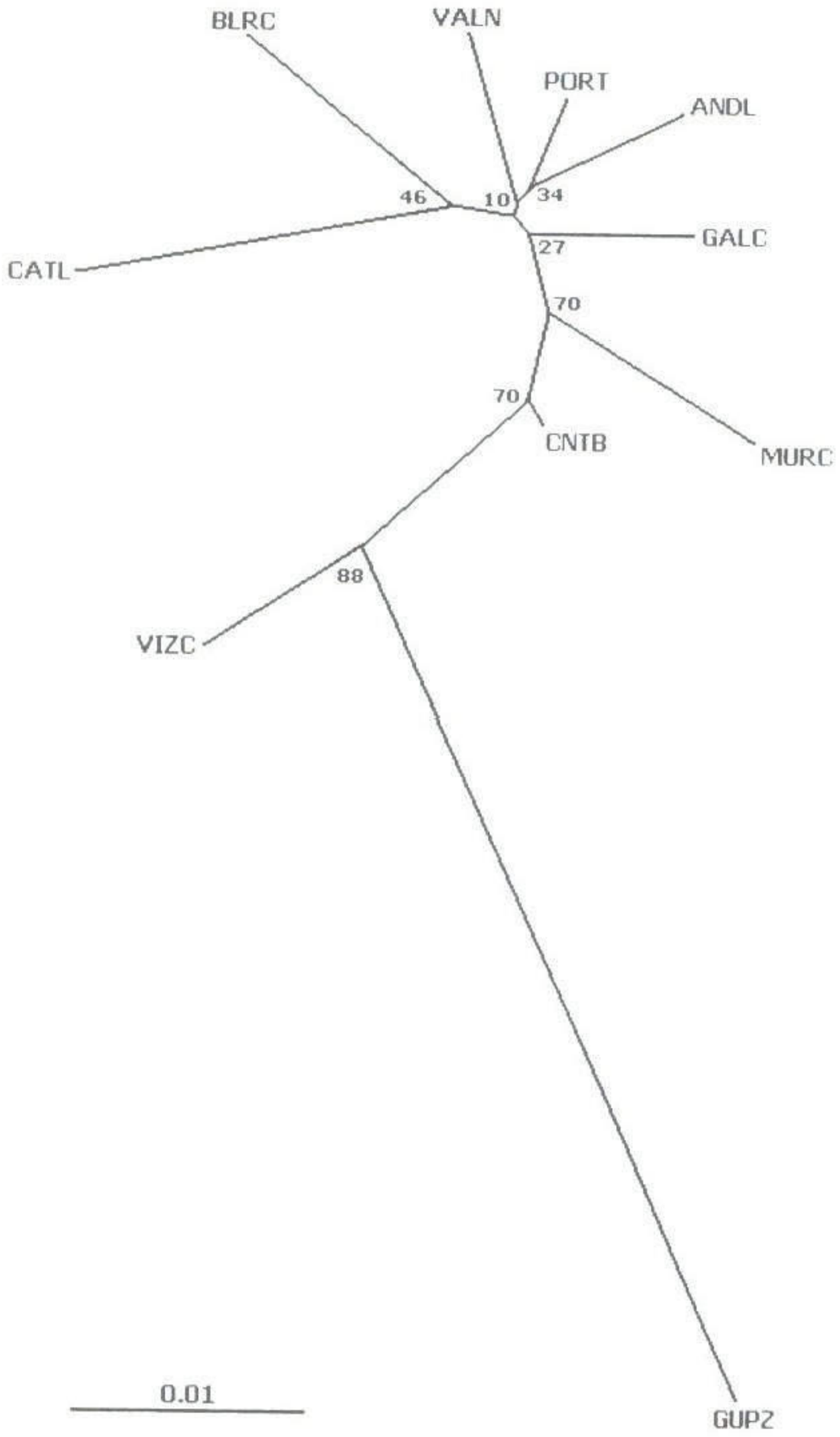

Figure 4. Unrooted neighbor-joining tree constructed from D_A genetic distances based on allele frequencies for nine STR loci in Iberian populations (1,000 bootstrap iterations). See Table 2 for population abbreviations.

Gene Flow Among Populations. To determine the relative amount of gene flow experienced by each population analyzed in this study, we compared the heterozygosity of each population against the genetic distance from the centroid (r_{ii}). The Guipuzcoa sample shows a lower than expected heterozygosity, relative to the theoretical regression line, suggesting a greater degree of isolation for this

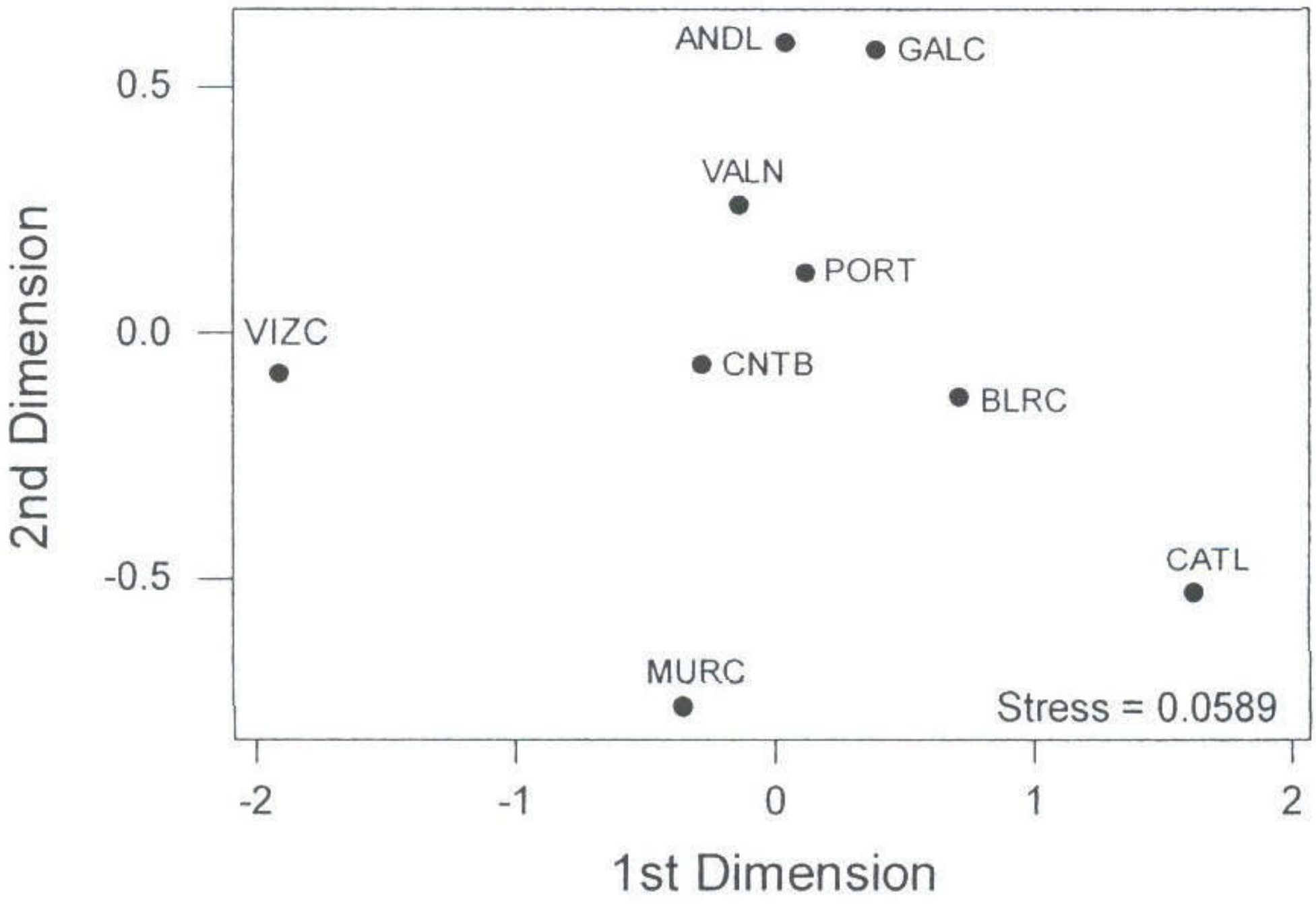

Figure 5. Multidimensional scaling applied to D_A genetic distances based on allele frequencies for nine STR loci in Iberian populations. See Table 2 for population abbreviations.

particular population. The Vizcayan Basques, on the other hand, are close to the expected heterozygosity value predicted by the regression model. Most of the comparative populations cluster around the theoretical regression line. The farthest removed populations are the Cantabrians, with a higher than expected heterozygosity, which indicates substantial gene flow, and the Maghreb sample, with a lower than expected value.

Discussion

In this study allele-frequency distributions of 13 STR loci are reported for a Basque sample from the Spanish province of Vizcaya. The values are generally within the range of those reported for other European populations (Pérez-Lezaun et al. 2000; Pepinski et al. 2001; Marjanovic et al. 2004). Heterozygosities for the individual loci are high (0.662–0.882) but are comparable to values of other commonly used STR markers in global populations (Jorde et al. 1997). Gene diversity analyses for nine of the selected STR loci reveal only a small proportion of the total diversity that is attributable to variability between European and North African populations, with an average G_{ST} score of only 0.7%. Interpopulational genetic distances, however, indicate that the Vizcayans, along with the Basque

sample from Guipuzcoa, are indeed genetic outliers relative to other neighboring groups for these nine STR loci.

Many different hypotheses have been proposed to account for the genetic distinctiveness of the Basque people. In particular, three models have been recently investigated and discussed by molecular anthropologists: (1) The Basque area was colonized by a long-distance migrating group of Neolithic agriculturalists from the Caucasus region (Calderón et al. 1998); (2) the Basques have paleo–North African origins (Arnaiz-Villena et al. 1999); and (3) the Basques are a relict population stemming from Mesolithic peoples that inhabited the Iberian peninsula (Bertranpetit and Cavalli-Sforza 1991). The first model has been supported by linguistic similarities between Euskera and North Caucasian languages (Gamkrelidze and Ivanov 1990) and the presence of immunoglobulin (*GM* and *KM*) allotypes with Central Asian affinities (Calderón et al. 1998, 2000). However, other genetic systems do not support this relationship (Bertorelle et al. 1995; Comas et al. 2000b; Semino et al. 2000; Sánchez-Velasco and Leyva-Cobián 2001) and archeological evidence does not indicate a population replacement in northern Spain during the Mesolithic-Neolithic transition (Zvelebil and Rowley-Conwy 1986; Jackes et al. 1997).

The second model is based on several population studies of *HLA* polymorphisms; these studies have found genetic ties between the Basques and certain North African groups (e.g., Algerians and Moroccans) (Martínez-Laso et al. 1995; Arnaiz-Villena et al. 1997). The results of the present study, however, fail to show any evidence of common ancestry or significant gene flow between these two regions. In the neighbor-joining tree based on pairwise D_A distances (Figure 2), the Maghreb, Moroccan, and Egyptian samples cluster together, whereas the Basques are positioned at the opposite end of the tree away from the North Africans with other European groups. The multidimensional-scaling plot (Figure 3) maintains this genetic orientation, with the North Africans clearly separated from the Europeans and Basques in two-dimensional space. Other studies based on classical genetic loci, autosomal STRs, *Alu* insertion polymorphisms, and Y-chromosome markers have corroborated this finding, underscoring the Gibraltar Strait as a strong barrier to genetic exchange (Bosch et al. 1997, 2000; Comas et al. 2000a; Flores et al. 2004).

In the neighbor-joining tree displayed in Figure 4 the two Basque samples cluster together at a robust 88% and are separated from the other Iberian groups by long phylogenetic branches. This pattern is also evident in the multidimensional-scaling plot (Figure 3), as both samples are removed from the Europeans and North Africans along the first dimension. The genetic differentiation of the Basques has been described in other population genetic studies and is commonly interpreted as evidence of a Basque ancestral relationship with Paleolithic and/or Mesolithic Iberians that avoided significant gene flow from expanding Neolithic agriculturalists throughout the Mediterranean basin (Cavalli-Sforza et al. 1994). However, the actual contribution of Near Eastern genes to the contemporary European and Iberian gene pools during the Neolithic transition has been a source

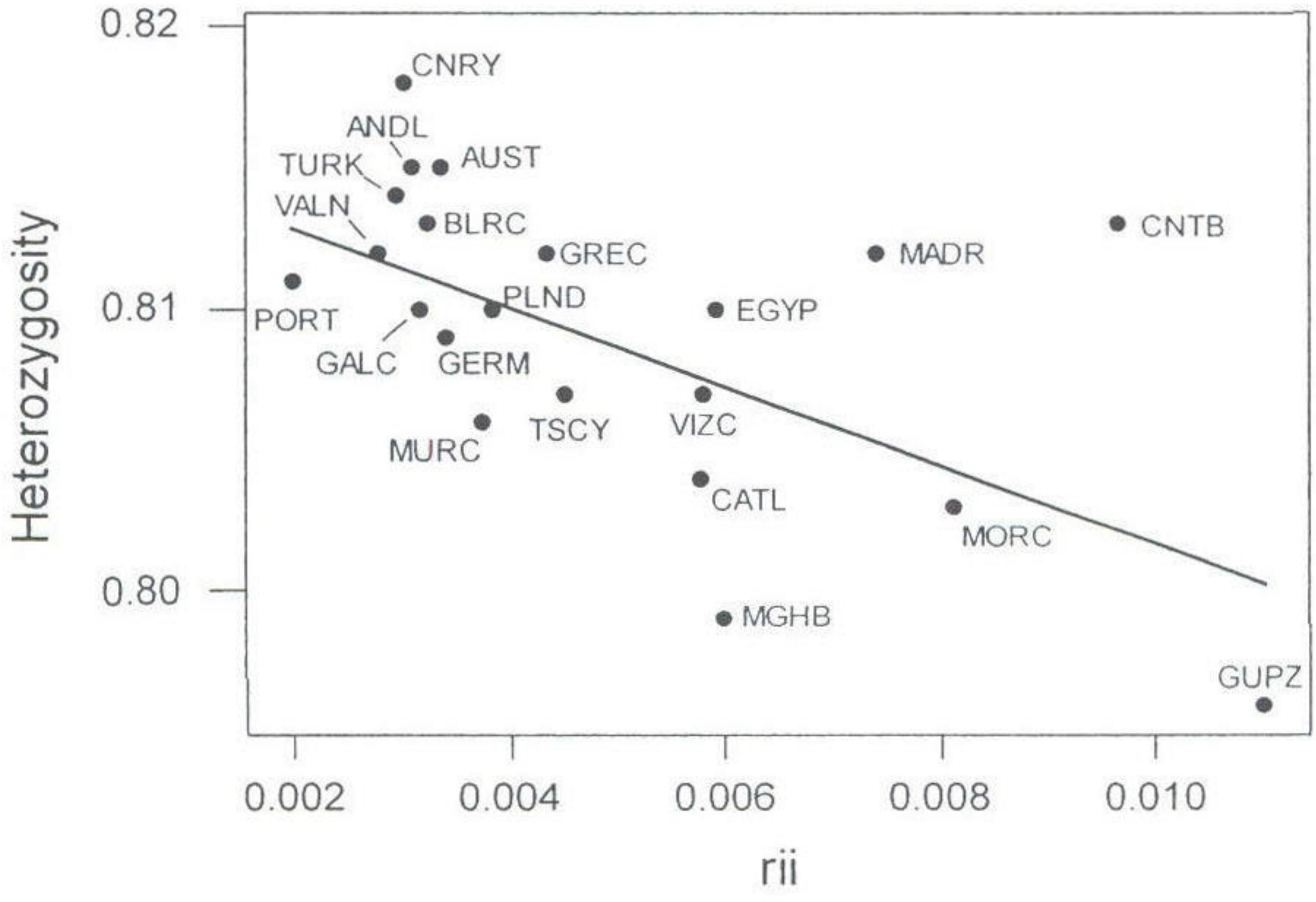

Figure 6. Regression plot of heterozygosity values and distances from the centroid for European and North African populations. See Table 2 for population abbreviations.

of much debate (Cavalli-Sforza and Minch 1997; Barbujani et al. 1998; Richards et al. 2000; Dupanloup et al. 2004), and thus whether or not the Basque differentiation identified in this study derives from a larger Paleolithic genetic component relative to other Europeans is unclear.

Although the Vizcayan and Guipuzcoan samples cluster in the neighbor-joining tree, the D_A genetic distance between the two is substantial, suggesting genetic heterogeneity within the Basque population. This is consistent with previous genetic studies that have identified significant population substructuring between Basque provinces and natural districts (Aguirre et al. 1991; Manzano et al. 1996, 2002; Iriondo et al. 2003). In the heterozygosity and centroid regression (Figure 6), the low heterozygosity and high r_{ii} value of the Guipuzcoans imply that this Basque population may have differentiated from other European populations as a result of genetic isolation and stochastic processes. But because this type of analysis does not provide information regarding the timing for periods of population isolation or gene flow, a clear explanation for this finding cannot be given. Conversely, the Vizcayan Basques exhibit values close to those predicted by the theoretical regression line, indicating a higher degree of gene flow and possibly reflecting the more urban character of this particular sample (i.e., Bilbao). However, this does not preclude the possibility that sampling technique may have played an important role in the low heterozygosity and high r_{ii} value

of the Guipuzcoan sample. As noted by Alonso et al. (2005), stringent criteria for selecting individuals of Basque ancestry can lead to reduced diversity and the impression of genetic isolation. In conclusion, the results of the present study reaffirm the genetic distinctiveness of the Basque population among both Europeans and North Africans and provide some evidence of population substructuring along provincial lines.

Acknowledgments We are grateful to all the Basque donors for participating in the present study. We thank Phil Melton, Alan Redd, and two anonymous reviewers for their critical reading of the manuscript and helpful critiques. This research was supported by the National Geographic Society through grant 6935-00.

Received 22 November 2005; revision received 19 April 2006.

Literature Cited

Aguirre, A. I., A. Vicario, L. I. Mazón et al. 1991. Are the Basques a single and unique population? *Am. J. Hum. Genet.* 49:450–458.

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

Amorim, A., L. Gusmão, and C. Alves. 2001. STR data (AmpFlSTR Profiler Plus) from north Portugal. *Forensic Sci. Int.* 115:119–121.

Anker, R., T. Steinbruck, and H. Donis-Keller. 1992. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. *Hum. Mol. Genet.* 1:137.

Anslinger, K., B. Rolf, and W. Keil. 2001. Evaluation and application of the AmpFlSTR Profiler Plus PCR amplification kit in a Bavarian population sample. *Int. J. Legal Med.* 114:278–280.

Arnaiz-Villena, A., J. Martínez-Laso, and J. Alonso-García. 1999. Iberia: Population genetics, anthropology, and linguistics. *Hum. Biol.* 71:725–743.

Arnaiz-Villena, A., J. Martínez-Laso, E. Gómez-Casado et al. 1997. Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by *HLA* allelic frequencies and haplotypes. *Immunogenetics* 47:37–43.

Asicioglu, F., F. Akyuz, U. Cetinkaya et al. 2002. Turkish population data on nine short tandem repeat loci: HumCSF1PO, HumTHO1, HumTPOX, HumFES/FPS, HumF13B, HumVWA, D3S1358, D7S820, D16S539. *Forensic Sci. Int.* 126:252–253.

Barbujani, G., G. Bertorelle, and L. Chikhi. 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. *Am. J. Hum. Genet.* 62:488–491.

Bertorelle, G., J. Bertranpetit, F. Calafell et al. 1995. Do Basque- and Caucasian-speaking populations share non-Indo-European ancestors? *Eur. J. Hum. Genet.* 3:256–263.

Bertranpetit, J., and L. L. Cavalli-Sforza. 1991. A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann. Hum. Genet.* 55:51–67.

Birren B., E. D. Green, and S. Klapholz, eds. 1997. *Genome Analysis: A Laboratory Manual*, v. 1. *Cold* Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Bosch, E., F. Calafell, A. Pérez-Lezaun et al. 1997. Population history of North Africa: Evidence from classical genetic markers. *Hum. Biol.* 69:295–311.

Bosch, E., F. Calafell, A. Pérez-Lezaun et al. 2000. Genetic structure of northwest Africa revealed by STR analysis. *Eur. J. Hum. Genet.* 8:360–366.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde et al. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457.

Boyd, W. C., and L. G. Boyd. 1937. New data on blood groups and other inherited factors in Europe and Egypt. *Am. J. Phys. Anthropol.* 23:49–70.

Calafell, F., and J. Bertranpetit. 1994. Principal component analysis of gene frequencies and the origin of Basques. *Am. J. Phys. Anthropol.* 93:201–215.

Calderón, R., A. Perez-Miranda, J. A. Peña et al. 2000. The genetic position of the autochthonous subpopulation of northern Navarre (Spain) in relation to other Basque subpopulations. A study based on *GM* and *KM* immunoglobulin allotypes. *Hum. Biol.* 72:619–640.

Calderón, R., C. Vidales, J. A. Peña et al. 1998. Immunoglobulin allotypes (*GM* and *KM*) in Basques from Spain: Approach to the origin of the Basque population. *Hum. Biol.* 72:619–640.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes.* Princeton, NJ: Princeton University Press.

Cavalli-Sforza, L. L., and E. Minch. 1997. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 61:247–251.

Chalmers, J. N. M., E. W. Ikin, and A. E. Mourant. 1949. The ABO, MN and Rh blood groups of the Basque people. *Am. J. Phys. Anthropol.* 7:529–544.

Comas, D., F. Calafell, N. Benchemsi et al. 2000a. *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: Evidence for a strong genetic boundary through the Gibraltar Straits. *Hum. Genet.* 107:312–319.

Comas, D., F. Calafell, N. Bendukidze et al. 2000b. Georgian and Kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am. J. Phys. Anthropol.* 112:5–16.

Côrte-Real, H. B., V. A. Macaulay, M. B. Richards et al. 1996. Genetic diversity in the Iberian peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.* 60:331–350.

Davrinche, C., C. Rivat, and L. Rivat-Peran. 1981. Allotypic variants of human C3 and properdin factor B in two French populations. *Blood Transfusion Immunohematol.* 24:309–318.

Dupanloup, I., G. Bertorelle, L. Chikhi et al. 2004. Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol. Biol. Evol.* 21:1361–1372.

Edwards, A., A. Citivello, H. A. Hammond et al. 1991. DNA typing and genetic mapping with trimeric tetrameric tandem repeats. *Am. J. Hum. Genet.* 49:746–756.

Edwards, A., H. A. Hammond, L. Jin et al. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241–253.

Esparza, B., C. Pestoni, M. D. Martin et al. 1995. Distribution of the *HLA-DQA1* polymorphism in the population of the Basque Country (Spain). *Gene Geogr.* 9:53–58.

Etcheverry, M. A. 1945. El factor Rhesus, su genética e importancia clinica. *Día Méd.* 17:1237–1251.

Farfán, M. J., V. Prieto, Y. Torres et al. 2001. STR data for the AmpFlSTR Profiler Plus and COfiler loci from the Maghreb (North Africa). *Forensic Sci. Int.* 121:199–200.

Felsenstein, J. 1993. *PHYLIP (Phylogeny Inference Package), Version 3.6.* Seattle, WA: University of Washington.

Fernandes, A. T., A. Brehm, C. Alves et al. 2002. Genetic profile of the Madeira archipelago population using the new PowerPlex16 System kit. *Forensic Sci. Int.* 125:281–283.

Flores, C., N. Maca-Meyer, A. M. González et al. 2004. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: Implications for population demography. *Eur. J. Hum. Genet.* 12:855–863.

Gamero, J. J., J. L. Romero, J. L. Gonzalez et al. 2000. A study on ten short tandem repeat systems: African immigrant and Spanish population data. *Forensic Sci. Int.* 110:167–177.

Gamkrelidze, T. V., and V. V. Ivanov. 1990. The early history of Indo-European languages. *Sci. Am.* 262:110–116.

Garcia, O., I. Uriarte, P. Martín et al. 2001. STR data from Basque country autochthonous population. *Forensic. Sci. Int.* 115:111–112.

Green, E. D., R. M. Mohr, J. R. Idol et al. 1991. Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: Application to the mapping of human chromosome 7 using yeast artificial chromosomes. *Genomics* 11:548–564.

Guo, S., and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.

Gusmão, L., P. Sánchez-Diz, C. Alves et al. 2000. Genetic diversity of nine STRs in two northwest Iberian populations: Galicia and northern Portugal. *Int. J. Legal Med.* 114:109–113.

Hammond, H. A., L. Jin, Y. Zhong et al. 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* 55:175–189.

Harpending, H. C., and R. H. Ward. 1982. Chemical systematics and human populations. In *Biochemical Aspects of Evolutionary Biology*, M. H. Nitechi, ed. Chicago: University of Chicago Press.

Hudson, T. J., L. D. Stein, S. S. Gerety et al. 1995. An STS-based map of the human genome. *Science* 270:1945–1954.

Iriondo, M., M. C. Barbero, and C. Manzano. 2003. DNA polymorphisms detect ancient barriers to gene flow in Basques. *Am. J. Phys. Anthropol.* 122:73–84.

Iriondo, M., C. Manzano, and C. de la Rúa. 1996. *HLA-DQA1* in autochthonous Basques: Description of a genocline for the *DQA1*0201* allele in Europe. *Int. J. Legal Med.* 109:181–185.

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

Jackes, M. D., D. Lubell, and C. Meiklejohn. 1997. On physical anthropological aspects of the Mesolithic-Neolithic transition in the Iberian Peninsula. *Curr. Anthropol.* 38:839–846.

Jauffrit, A., H. El Amri, F. Airaud et al. 2003. DNA short tandem repeat profiling of Morocco. *J. Forensic Sci.* 48:458–459.

Jorde, L. B., A. R. Rogers, M. Bamshad et al. 1997. Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci. USA* 94:3100–3103.

Kimpton, C., A. Walton, and P. Gill. 1992. A further tetranucleotide repeat polymorphism in the vWF gene. *Hum. Mol. Genet.* 1:287.

Klintschar, M., N. Al-Hammadi, and B. Reichenpfader. 1999. Population genetic studies on the tetrameric short tandem repeat loci *D3S1358*, *VWA*, *FGA*, *D8S1179*, *D21S11*, *D5S818*, *D13S317*, and *D7S820* in Egypt. *Forensic Sci. Int.* 104:23–31.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.

Li, C. C. 1976. *First Course in Population Genetics*. Pacific Grove, CA: Boxwood.

Li, H., L. Schmidt, M. H. Wei et al. 1993. Three tetranucleotide polymorphisms for loci: *D3S13502*, *D3S1358*, *D3S1359*. *Hum. Mol. Genet.* 2:1327.

Manzano, C., A. I. Aguirre, M. Iriondo et al. 1996. Genetic polymorphisms of the Basques from Gipuzkoa: Genetic heterogeneity of the Basque population. *Ann. Hum. Biol.* 23:285–296.

Manzano, C., C. de la Rúa, M. Iriondo et al. 2002. Structuring the genetic heterogeneity of the Basque population: A view from classical polymorphisms. *Hum. Biol.* 74:51–74.

Manzano, C., P. Moral, C. de la Rúa et al. 1993. Serum protein polymorphisms (GC, TF, and PI subtypes) in the Basque population of Alava. *Hum. Hered.* 43:121–125.

Marjanovic, D., L. Kapur, K. Drobnic et al. 2004. Comparative study of genetic variation at 15 STR loci in three isolated populations of the Bosnian mountain area. *Hum. Biol.* 76:15–31.

Martínez-Laso, J., D. De Juan, N. Martínez-Quiles et al. 1995. The contribution of the HLA-A-B-C and -DR-DQ DNA typing to the study of the origins of Spaniards and Basques. *Tissue Antigens* 45:237–245.

Mills, K. A., D. Even, and J. C. Murray. 1992. Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (*FGA*). *Hum. Mol. Genet.* 1:779.

Nei, M. 1973. Analysis of gene diversity in subdivided population. *Proc. Natl. Acad. Sci. USA* 70:3321–3323.

Nei, M. 1987. *Molecular Evolutionary Genetics.* New York: Columbia University Press.

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19:153–170.

Neuhuber, F., M. Radacher, N. Meisner et al. 1999. Nine STR markers plus amelogenin (AmpFlSTR Profiler Plus): A forensic study in an Austrian population. *Int. J. Legal Med.* 113:60–62.

Oldroyd, N. J., A. Urquhart, C. P. Kimpton et al. 1995. A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification. *Electrophoresis* 16:334–337.

Ota, T. 1993. *DISPAN: A Software for Genetic Distances and Neighbor-Joining Calculations.* University Park, PA: Pennsylvania State University.

Peña, J. A., R. Calderón, A. Perez-Miranda et al. 2002. Microsatellite DNA markers from *HLA* region (*D6S105*, *D6S265* and *TNFa*) in autochthonous Basques from northern Navarre (Spain). *Ann. Hum. Biol.* 29:176–191.

Pepinski, W., J. Janica, M. Skawronska et al. 2001. Population genetics of 15 STR loci in the population of Podlasie (NE Poland). *Forensic Sci. Int.* 124:226–227.

Pérez-Lezaun, A., F. Calafell, J. Clarimón et al. 2000. Allele frequencies of 13 short tandem repeats in population samples from the Iberian peninsula and Northern Africa. *Int. J. Legal Med.* 113:208–214.

Pérez-Lezaun, A., F. Calafell, E. Mateu et al. 1997. Microsatellite variation and the differentiation of modern humans. *Hum. Genet.* 99:1–7.

Polymeropoulos, M. H., D. S. Rath, H. Xiao et al. 1991. Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nucleic Acids Res.* 19:3753.

Ricci, U., I. Sani, L. Giunti et al. 2002. Analysis of 13 tetrameric short tandem repeat loci in a population of Tuscany (central Italy) performed by means of an automated infrared sequencer. *Forensic Sci. Int.* 125:83–85.

Richards, M., H. Côrte-Real, P. Forster et al. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 59:185–203.

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

Rohlf, F. J. 2002. *NTSYSpc: Numerical Taxonomy System, ver. 2.1.* Setauket, NY: Exeter Publishing.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.

Sánchez-Diz, P., M. V. Lareu, M. Brion et al. 2002. STR data for the AmpFlSTR Profiler Plus loci from Greece. *Forensic Sci. Int.* 126:265–266.

Sánchez-Velasco, P., and F. Leyva-Cobián. 2001. The *HLA* class I and class II allele frequencies studied at the DNA level in the Svanetian population (Upper Caucasus) and their relationships to Western European populations. *Tissue Antigens* 58:223–233.

Sanz, P., V. Prieto, I. Flores et al. 2001. Population data of 13 STRs in southern Spain (Andalucía). *Forensic Sci. Int.* 119:113–115.

Schneider, S., D. Roessli, and L. Excoffier. 2000. *Arlequin, ver. 2.00: A Software for Population Genetic Analysis.* Geneva: Genetics and Biometry Laboratory, University of Geneva.

Semino, O., G. Passarino, P. J. Oefner et al. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y-chromosome perspective. *Science* 290:1155–1159.

Sharma, V., and M. Litt. 1992. Tetranucleotide repeat polymorphism at the *D21S11* locus. *Hum. Mol. Genet.* 1:67.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399.

Tomás, C., A. Picornell, J. A. Castro et al. 2000. Genetic variability at nine STR loci in the Chueta (Majorcan Jews) and the Balearic populations investigated by a single multiplex reaction. *Int. J. Legal Med.* 113:263–267.

Tomás, C., A. Picornell, J. A. Castro et al. 2001. STR data for the AmpFlSTR profiler plus loci from Majorcan, Minorcan, and Valencian populations (eastern Spain). *Forensic Sci. Int.* 121:201–204.

Toria, A., M. R. Moya-Quiles, A. Luna et al. 2002. STR data for the AmpFlSTR Profiler Plus loci from Murcia region (SE Spain). *Forensic Sci. Int.* 129:140–141.

Torroni, A., H. J. Bandelt, L. D'Urbano et al. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* 62:1137–1152.

Trask, L. 1997. *The History of Basque.* London: Routledge.

Urquhart, A., N. J. Oldroyd, C. P. Kimpton et al. 1995. Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *Biotechniques* 18:116–121.

Vergnes, H. J., J. Sevin, and M. Peralta. 1980. Plasma cholinesterase variants in population samples from southwest France. *Ann. Hum. Biol.* 8:59–63.

Zarrabeitia, M. T., and J. A. Riancho. 2001. Population data on nine STRs from Cantabria, a mountainous region in northern Spain. *Forensic Sci. Int.* 122:175–177.

Zvelebil, M., and P. Rowley-Conwy, eds. 1986. *Hunters in Transition: Mesolithic Societies of Temperate Eurasia and Their Transition to Farming.* Cambridge, England: Cambridge University Press.