ORIGINAL ARTICLE

# Polymorphic *Alu* insertions and the genetic structure of Iberian Basques

**S. García-Obregón · M. A. Alfonso-Sánchez · A. M. Pérez-Miranda · M. M. de Pancorbo · J. A. Peña**

**Abstract** Eight *Alu* sequences (ACE, TPA25, PV92, APO, FXIIIB, D1, A25 and B65) were analyzed in two samples from Navarre and Guipúzcoa provinces (Basque Country, Spain). *Alu* data for other European, Caucasus and North African populations were compiled from the literature for comparison purposes to assess the genetic relationships of the Basques in a broader geographic context. Results of both MDS plot and AMOVA revealed spatial heterogeneity among these three population clusters clearly defined by geography. On the contrary, no substantial genetic heterogeneity was found between the Basque samples, or between Basques and other Europeans (excluding Caucasus populations). Moreover, the genetic information obtained from *Alu* data conflicts with hypotheses linking the origin of Basques with populations from North Africa (Berbers) or from the Caucasus region (Georgia). In order to explain the reduced genetic heterogeneity detected by *Alu* insertions among Basque subpopulations, values of the Wright's $F_{ST}$ statistic were estimated for both *Alu* markers and a set of short tandem repeats (STRs) in terms of two geographical scales: (1) the Basque Country, (2) Europe (including Basques). In the Basque area, estimates of Wahlund's effect for both genetic markers showed no statistical difference between Basque subpopulations. However, when this analysis was performed on a European scale, $F_{ST}$ values were significantly higher for *Alu* insertions than for STR alleles. From these results, we suggest that the spatial heterogeneity of the Basque gene pool identified in previous polymorphism studies is relatively recent and probably caused by a differential process of genetic admixture with non-Basque neighboring populations modulated by the effect of a linguistic barrier to random mating.

**Keywords** *Alu* insertions · Genetic heterogeneity · Gene flow · Wright's $F_{ST}$ · Basques

S. García-Obregón · M. A. Alfonso-Sánchez · A. M. Pérez-Miranda · J. A. Peña (✉)
Departamento de Genética,
Antropología Física y Fisiología Animal,
Facultad de Ciencia y Tecnología,
Universidad del País Vasco, Apartado 644,
48080 Bilbao, Spain
e-mail: joseangel.pena@ehu.es

M. M. de Pancorbo
Departamento de Zoología y Dinámica Celular Animal,
Facultad de Farmacia, Universidad del Pais Vasco,
48080 Bilbao, Spain

## Introduction

The systematic study of different types of molecular polymorphisms in a human population can provide detailed information on the demographic and micro-evolutionary processes that have affected it over time. Such polymorphisms include a number of DNA markers located in the non-recombining region of the Y-chromosome (NRY) with the potential to provide information on male-specific patterns of migration in the past, namely Y-chromosomal short tandem repeats and single nucleotide polymorphisms (Y-STRs, Y-SNPs), whereas analyses of the mitochondrial DNA (mtDNA) can assist in the clarification female-mediated migration episodes. Still other polymorphisms are useful for describing the joint evolution of the maternal and paternal lineages (Jorde et al. 1995; Lell and Wallace 2000; Richards 2003). At a different level,

there are molecular markers that have changed little throughout evolution, such as *Alu* elements and SNPs, whereas other markers stand out as having high mutation rates, such as STRs (Jorde et al. 1997; Watkins et al. 2001, 2003). The more conservative of these polymorphisms are considered to be suitable markers for studying population phylogenetics due to their abundance in the genome and their low mutation rates. The less conservative, hypervariable DNA markers reveal data on the most recent micro-evolutionary changes undergone by a given population.

*Alu* elements represent around 10% of the human genome, with around 1,400,000 copies distributed throughout. These insertions are part of the most abundant short interspersed elements (SINEs) in the human genome, and they are defined as sequences of approximately 300 base pairs (bp) in length ancestrally originating from the 7SL RNA gene by retro-transposition (reviewed in Batzer and Deininger 2002). Some interesting characteristics of the *Alu* insertions make them valuable markers in phylogenetic analyses. The main feature of *Alu* polymorphism is stability; because alleles are identical by descent, it is highly unlikely that the same *Alu* insertion could occur more than once independently at the same locus. This means that polymorphic *Alu* insertions reflect, in general unique evolutionary events. Furthermore, the absence of the insertion facilitates a knowledge of the ancestral state, which is an advantageous attribute in investigations focused on increasingly disentangling the demographic, genetic and evolutionary history of the human species (Batzer et al. 1994).

It is possible that the study of *Alu* insertions would shed some light on human evolutionary genetic processes in the European and Mediterranean context that are currently being debated among specialists. These include (1) the origin of Basques and the genetic heterogeneity of the present-day Basque population; (2) the impact of North African migrations on the genetic background of the Iberian populations; (3) the peopling of Europe. The choice of polymorphisms with a suitable level of resolution and their analysis in a large enough series of representative samples from different geographic regions should provide interesting data to help clarifying these questions. Therefore, one of the main aims of this study, which is a continuation of previous investigations carried out by the authors (de Pancorbo et al. 2001; Peña et al. 2002; Pérez-Miranda et al. 2003, 2004, 2005; Alfonso-Sánchez et al. 2006; García-Obregón et al. 2006) was to create a robust genetic database which may help in the gradual reconstruction of the peopling processes of Europe and the Mediterranean Basin.

In this study, we have genetically characterized two samples of Basques based on eight polymorphic *Alu* insertions (ACE, TPA25, PV92, APO, FXIIIB, D1, A25 and B65), which were selected bearing in mind the availability of databases for comparisons. Samples were collected from the autochthonous population settled in Navarre and Guipúzcoa provinces, both of which are located in the historical Basque territory (Northern Spain). These areas were selected for two basic reasons: (1) Northern Navarre and Guipúzcoa are the two Basque regions where the Basque language (Euskera) has traditionally been more deep-rooted (Alfonso-Sánchez et al. 2005); (2) Navarre is the only Basque region for which *Alu* insertion frequencies are not available. The findings on *Alu* diversity in these native Basque populations were then examined in a broader geographic context. To that end, *Alu* data for European, Caucasus and North African populations were compiled from the literature. With this integrative approach we sought to analyze the degree of genetic heterogeneity among geographical subpopulations of the Basque Country as well as to assess the genetic relationships of the autochthonous Basques with North African and Caucasus populations, since both of the latter population groups have been linked to the origin of Basques in some previous works (Arnaiz-Villena et al. 1997; Calderón et al. 1998). In addition, *Alu* data presented herein may be useful in the interpretation of gene flow processes that have contributed to the shaping of the European gene pool.

## Material and methods

The Basque area is located at the western end of the Pyrenees on the Bay of Biscay area of the Atlantic Coast, astride the border between France and Spain. For further details on the geographic, demographic and linguistic characteristics of the study area (Guipúzcoa and Navarre provinces), the reader is referred to previous publications (Pérez-Miranda et al. 2003, 2005).

Whole blood samples were collected in EDTA vacutainer tubes by venipuncture from unrelated healthy individuals living in Guipúzcoa province ($n = 94$) and in the northern fringe of Navarre province ($n = 109$). Only autochthonous (native) Basque individuals were included in the present analysis. In this study, Basque surnames and birthplaces of individuals and ancestors (recorded back to the third generation) were the criteria employed to define local autochthony; therefore, all donors were interviewed to obtain information on the geographical origins of their parents and grandparents. Ethical guidelines for research

with humans were adhered to as stipulated by the Ethical Committee of the University of the Basque Country. All blood donors gave their informed consent prior to inclusion in the sample.

Genomic DNA was extracted from peripheral blood using the standard phenol-chloroform procedure and stored at –20°C. Eight autosomal *Alu* insertions (ACE, TPA25, PV92, APO, FXIIIB, D1, A25 and B65) were genotyped in both samples. Additional information on the PCR amplification conditions and agarose gel electrophoresis can be found in García-Obregón et al. (2006).

### Statistical analyses

Allelic frequencies for the eight *Alu* loci typed in the Navarre and Guipúzcoa collections were calculated by direct counting. Gene diversity (GD) was computed using the POWER MARKER v. 3.0 program (Liu and Muse 2004). To test for Hardy-Weinberg equilibrium (HWE), we carried out a Fisher's exact probability test to estimate *P* values (Guo and Thompson 1992) using ARLEQUIN v. 3.0 (Excoffier et al. 2005).

The genetic associations and population affinities of Navarre and Guipúzcoa samples in a broader geographic context were analyzed by compiling *Alu* data on European, Caucasus and North African populations from previously published works (Stoneking et al. 1997; Comas et al. 2000, 2004; Nasidze et al. 2001; Romualdi et al. 2002; Maca-Meyer et al. 2004; García-Obregón et al. 2006). This genetic information was employed to compute $F_{ST}$ unbiased genetic distances (Reynolds et al. 1983) between all pairs of populations. To represent the resultant $F_{ST}$ genetic distance matrix in a two-dimensional space, nonmetric multidimensional scaling (MDS) analysis (Kruskal 1964) was performed using the SPSS (ver. 13.0; SPSS, Chicago, Ill.) statistical package. The association between geographic and genetic distances for the populations of the Basque Country was analyzed by constructing a consensus topogenetic map. To that end, a nonmetric MDS was first performed to represent the genetic distance matrix. Later, the first two eigenvectors of the MDS plot obtained were rotated to maximum congruity with the geographic coordinates, using the methodology described by Lalouel (1973). In order to ascertain the proportions of the genetic variance due to differences within and between populations, genetic variance was hierarchically apportioned according to geographic criteria through the analysis of molecular variance (AMOVA) (Excoffier et al. 2005) using the ARLEQUIN program. In this statistical analysis, a permutation procedure allows testing the significance of

the fixation indices $F_{CT}$, $F_{SC}$ and $F_{ST}$ that measure the relative contribution of the genetic variation between groups, between populations within groups, and within populations, respectively. We subsequently established an overall test (including all *Alu* loci) in order to check the statistical significance of $F_{CT}$ values by combining the separate probability values for each locus through the equation

$$\chi^2_{[2k]} = -2\sum_{I=1}^{k} \ln p_I$$

where *k* is the number of loci and $p_I$ the separate probability value associated with the $F_{CT}$ values for each *I* locus (Sokal and Rohlf 1997).

Finally, with the aim of interpreting the genetic heterogeneity observed among Basque subpopulations and assessing the Wahlund's effect, Wright's $F_{ST}$ statistic values were estimated from both *Alu* insertions and a series of forensic STRs. To that end, allelic frequencies for several European and North-African samples compiled in a previous publication (Pérez-Miranda et al. 2005) were considered. The statistical significance of the differences in Wright's $F_{ST}$ values obtained for both types of genetic markers was then assessed using the Mann-Whitney nonparametric *U* test.

## Results

The *Alu* insertion frequencies for the eight loci typed in the study populations are listed in Table 1. As can be noted, all *Alu* insertions were polymorphic in both populations, with APO being the closest to fixation. This locus showed insertion frequencies of 0.950 in Guipúzcoa and 0.970 in Navarre. Conversely, A25 was the least frequent element, at just over 0.124 in Navarre and 0.156 in Guipúzcoa. For the remaining *Alu* markers, the allelic frequencies oscillated between 0.239 (PV92) and 0.557 (B65) in Guipúzcoa, and between 0.154 (PV92) and 0.592 (TPA25) in Navarre. HWE was assessed by an exact test to calculate the *P* value using the Markov-chain Monte Carlo method (Guo and Thompson 1992). No significant departure from HWE expectations was detected in most of the eight *Alu* insertions analyzed, with the exceptions of ACE in Navarre and B65 in Guipúzcoa.

*Alu* insertions are classed as biallelic markers according to the origin of their polymorphism. Interestingly, the analysis of the PV92 locus revealed three heterozygotic individuals for a third allele in the sample from Guipúzcoa province. This variant consisted of

**Table 1** *Alu* insertion frequencies with their standard errors (±SE) and gene diversity (GD) in Basque samples from Guipúzcoa and Navarre provinces (Spain)

| | Population | | | | | |
|---|---|---|---|---|---|---|
| | Guipúzcoa | | | Navarre | | |
| *Alu* locus | 2N[a] | Frequency ±SE | GD | 2N[a] | Frequency ±SE | GD |
| ACE | 188 | 0.425 ± 0.033 | 0.489 | 206 | 0.320 ± 0.027 | 0.435 |
| TPA25 | 188 | 0.553 ± 0.034 | 0.494 | 218 | 0.592 ± 0.034 | 0.483 |
| PV92 | 188 | 0.239 ± 0.0387[b] | 0.371 | 208 | 0.154 ± 0.026 | 0.260 |
| APO | 182 | 0.950 ± 0.017 | 0.094 | 202 | 0.970 ± 0.011 | 0.058 |
| FXIIIB | 188 | 0.441 ± 0.036 | 0.493 | 208 | 0.490 ± 0.034 | 0.500 |
| D1 | 186 | 0.473 ± 0.034 | 0.498 | 204 | 0.372 ± 0.032 | 0.467 |
| A25 | 186 | 0.156 ± 0.025 | 0.263 | 202 | 0.124 ± 0.021 | 0.217 |
| B65 | 176 | 0.557 ± 0.032 | 0.493 | 208 | 0.466 ± 0.035 | 0.498 |

[a] 2N, Sample size in number of chromosomes analyzed

[b] Three chromosomes with a second insertion are included

a second *Alu* insertion in an existing *Alu* element. To date, this uncommon allele has been described only in Basque and North African populations (Comas et al. 2001).

The degree of genetic variability in the Basque samples was assessed by computing the GD for each locus (see Table 1). As expected, the lowest figures of GD were observed in the *Alu* loci closest to fixation (APO), both in Guipúzcoa (0.094) and in Navarre (0.058). On the contrary, the highest GD values were found in those *Alu* markers showing insertion frequencies of around 0.50. These were the cases of D1 in Guipúzcoa (0.498) and of FXIIIB in Navarre (0.500).

For a thorough analysis of the genetic structure of Basques, data on *Alu* insertions from previous publications were compiled. This database included a sample of native Basques with non-specified origin, from now on referred to as "general" Basques (Comas et al. 2000). We also included the autochthonous samples studied in the article by de Pancorbo et al. (2001), namely Basques from Biscay, Alava and Guipúzcoa, in addition to a sample from the resident population of the Basque Country (not all individuals are, necessarily, native people). Because *Alu* markers A25 and B65 were not genotyped in most of the cited samples, these two insertions were not considered in our analysis. Although the exact test of population differentiation (data not shown) failed to detect significant genetic heterogeneity for the whole set of Basque samples, some significant differences were found for several *Alu* loci in the following pairs of populations: Alava and Navarre (ACE, $P = 0.036$), Guipúzcoa (this study) and Biscay (D1, $P = 0.007$) and Navarre and the resident population (APO, $P = 0.035$).

A distance matrix (Reynolds' $F_{ST}$) was computed and represented in a bidimensional space by means of

a nonmetric MDS plot. Subsequently, the first two eigenvectors were fitted up to maximum congruity with the geographical coordinates of the samples' origin (Fig. 1). In this analysis, resident Basques and general Basques were represented as witness populations, since it is not possible to assign specific geographical coordinates to these samples. Resident Basques displayed the more deviated genetic coordinates with respect to
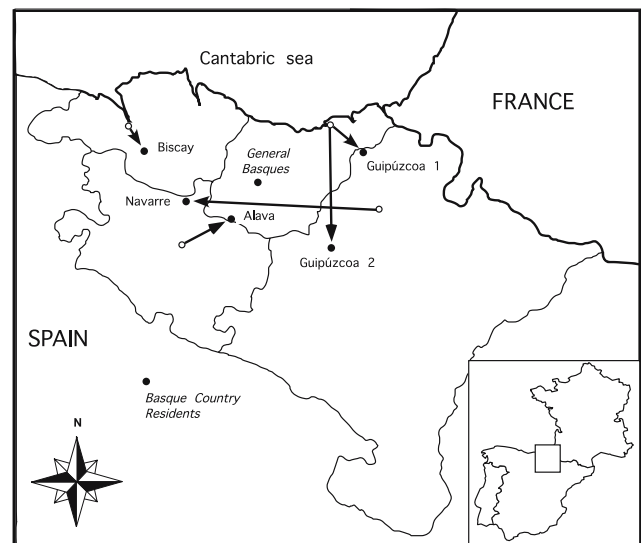


**Fig. 1** Matrix fitting of geographic and genetic coordinates for six Basque subpopulations. *Open circles* represent geographic locations of the populations analyzed, *arrows* indicate the location predicted by genetic kinship (*full circles*). Study populations are Navarre and Guipúzcoa 2. Data sources: Guipúzcoa 1, Biscay, Alava and Basque Country Residents – de Pancorbo et al. (2001); general Basques – Comas et al. (2000). Because it is not possible to exactly determine the geographical coordinates for the samples of general Basques (*Basques*) and resident Basques (*Basque Country Residents*), only the genetic coordinates are represented in both cases

the rest of Basque subpopulations. This is mostly the consequence of a relatively high level of genetic admixture, bearing in mind that the origin of many of the individuals who compose this sample is located elsewhere in Iberia (de Pancorbo et al. 2001). In a previous work, these authors estimated the degree of genetic admixture of the present-day resident population of the Basque Country and found a contribution from the (non-Basque) Iberian populations to the Basque gene pool of around 44% (Peña et al. 2006). On the other hand, genetic coordinates obtained for the Biscay, Navarre and Alava samples tend to be closer than expected according to their geographic distribution. These subpopulations occupy peripheral positions within the traditional Basque territories, bordering on other Spanish regions where Romance languages have predominated for centuries. Conversely, genetic coordinates of the two Guipúzcoa samples showed a certain tendency to drift apart from the rest of the Basque collections. Guipúzcoa is the only Basque "historical territory" that is surrounded by other Basque provinces (Biscay, Alava, Navarre and the French Basque Country) and, therefore, the least influenced by gene flow processes 'imported' from the neighboring Spanish-speaking provinces. Finally, the genetic location of general Basques suggests an intermediate zone on the topogenetic map; this was expected taking into account that this sample cannot be ascribed to any particular Basque area. Ultimately, general Basques can be seen as the representation of the average Basque gene pool.

For comparison purposes, *Alu* data from other European, Caucasus and North African populations were compiled from the literature. In this analysis, only those samples with data available for the eight *Alu* insertions were included, whereas populations with very small sample sizes were excluded. Geographic associations and/or genetic affinities among populations were analyzed by computing $F_{ST}$ genetic distances between pairs of samples and representing them in two-dimensional spaces by means of nonmetric MDS. The MDS plot accounted for 99.3% of the total variance, with a coefficient of stress of 0.053 (Fig. 2). Consistent with earlier human diversity studies based on *Alu* elements (Stoneking et al. 1997; Watkins et al. 2001, García-Obregón et al. 2006), European, Caucasus and North African populations formed three visibly separated clusters.

Firstly, the Caucasus populations can be seen to display a very high genetic heterogeneity despite their limited geographical distribution, as can be inferred from their remarkable dispersion along both dimension I and dimension II of the MDS plot. In particular, there
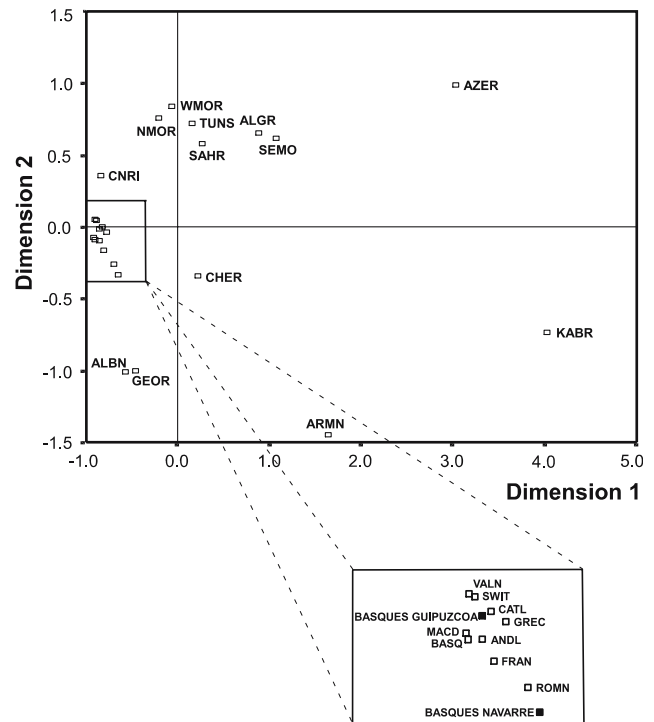


**Fig. 2** Nonmetric multidimensional scaling (MDS) applied on $F_{ST}$ genetic distance matrix. Genetic distances were computed from allelic frequencies of eight *Alu* insertions to analyze genetic relationships among 24 European, North African and Caucasus populations. *ANDL* Andalusia, *ALBN* Albania, *ALGR* Algeria, *ARMN* Armenia, *AZER* Azerbaijan, *BASQ* Basque Country, *CATL* Catalonia, *CHER* Cherkesia, *CNRI* Canary Islands, *FRAN* France, *GEOR* Georgia, *KABR* Kabardinia, *MACD* Macedonia, *GREC* Greece, *NMOR* North Morocco, *SEMO* Southeast Morocco, *WMOR* West Morocco, *ROMN* Romania, *SAHR* Sahara, *SWIT* Switzerland, *TUNS* Tunisia, *VALN* Valencia. Study populations are Basques Guipúzcoa and Basques Navarre

were conspicuous separations between the plots of the collections from Kabardinia (KABR), Azerbaijan (AZER) and Armenia (ARMN); these collections were also markedly distant from the rest of the populations included in the analysis. On the other hand, Cherkessians (CHER) and Georgians (GEOR) grouped closer to the European cluster. The North African cluster seems to show only a moderate within-group genetic heterogeneity. These samples grouped mostly in the upper-right quadrant delimited by the positive segments of both dimension I and II and visibly segregated from the Caucasus and European populations. A third major grouping consists of the European populations not located in the Caucasus; this grouping is concentrated very close to the centroid of the bidimensional representation and constitutes a very homogeneous cluster from a genetic standpoint. Two populations are worth highlighting within this group: Albania (ALBN), with *Alu* insertion frequencies sim-

ilar than those estimated in GEOR, and the Canary Isles (CNRI), whose position is in agreement with historical origins and with a geography positioned in an interim point between European and North African samples. More specifically, the Basque samples were dispersed within the European cluster, staying distant from both the Caucasus and North African populations.

Because findings of the MDS plot reveal an overlapping of *Alu* frequencies with the geographical distribution of populations, we analyzed whether the genetic heterogeneity detected is spatially structured by hierarchical apportioning of the genetic diversity by carrying out an analysis of the molecular variance (AMOVA). The populations considered in this study were grouped according to geographical criteria: North African, European and Caucasus populations. Upon assignment of the populations within these three broad geographic regions, AMOVA analyses were performed for each of the *Alu* insertions examined (Table 2), and the results revealed statistically significant genetic heterogeneity among the three population groups ($F_{CT}$) for all of the *Alu* loci, with the exception of B65 (0.07% of total variation accounted for, ns). The overall AMOVA test (Sokal and Rohlf 1997) performed by considering all *Alu* elements jointly also yielded highly significant differences among groups ($P < 0.001$). In addition, the heterogeneity among populations within groups ($F_{SC}$) proved to be signifi-

cant in four *Alu* markers, namely ACE (0.51%, $P < 0.01$), FXIIIB (1.91%, $P < 0.01$), D1 (0.90%, $P < 0.01$) and B65 (0.64%, $P < 0.05$), with the overall test being significant ($P < 0.001$) accordingly. This microgeographic differentiation of the *Alu* frequencies was mostly conditioned by the Caucasus group, in which significant interpopulation differences were found for FXIIIB (17.5%, $P < 0.001$), B65 (6.6%, $P < 0.001$), D1 (8.0%, $P < 0.01$) and PV92 (9.2%, $P < 0.001$), thus reflecting the high dispersion of these samples in the MDS plot. As regards to the European cluster, spatial structuring of the *Alu* frequencies was only found for ACE locus (0.7%, $P < 0.05$), whereas no statistical difference was detected among North African populations.

In order to explain the scarce genetic heterogeneity detected by the *Alu* insertions among subpopulations of the Basque Country, values of the Wright's $F_{ST}$ statistic were estimated for the *Alu* markers and compared with the obtained ones in a set of 13 forensic autosomal microsatellites or STRs for the same populations (data from Pérez-Miranda et al. 2005). Peña et al. (1997) demonstrated that Wright's $F_{ST}$ statistic depends on the number of subpopulations and their effective size (Ne). Thus, it is essential to consider the same group of subpopulations so that the $F_{ST}$ values obtained for different genetic loci are comparable. Comparisons of Wright's $F_{ST}$ values between *Alu* elements and STR loci were performed at two different geographic scales: (1) considering only subpopulations of the Basque Country, (2) including other European samples in addition to the Basque subpopulations. As can be noticed in Fig. 3, *Alu* insertions possess average frequencies generally higher than those observed for STR alleles in the Basque Country. However, taking the Basque population on a whole, the estimates of Wahlund's effect ($F_{ST}$) for both types of genetic markers showed no statistical difference, as verified by a Mann-Whitney $U$ test ($U = 385.5$, ns). Conversely, when this analysis was carried out for a wider group of populations by considering data on the same polymorphisms (*Alu* and STRs) in the European samples, the results of the nonparametric $U$ test indicated that Wright's $F_{ST}$ values are significantly higher for *Alu* insertions than for STR alleles ($U = 44.0$, $P < 0.001$) (see Fig. 3). In view of these results, it can be concluded that the heterogeneity of the *Alu* insertions is lower in the Basque samples than in Europe as a whole (taking a set of autosomal STRs as reference). This finding suggests that the processes that have promoted, for the most part, the genetic microdifferentiation in both geographical contexts should not have been equivalent.

**Table 2** Results of the analysis of molecular variance ($F_{ST}$, $F_{SC}$ and $F_{CT}$) for eight *Alu* loci three different population clusters[a] classified according to geography

| *Alu* marker | $F_{CT}$[b] (%) | $F_{SC}$[c] (%) | $F_{ST}$[d] (%) |
|---|---|---|---|
| ACE | 1.06** | 0.51** | 98.43* |
| TPA 25 | 1.38** | 0.00 NS | 98.73 NS |
| PV92 | 2.57** | 0.37 NS | 97.07** |
| APO | 2.64** | 0.05 NS | 97.31** |
| FXIIIB | 1.68* | 1.91** | 96.41** |
| D1 | 1.69** | 0.90** | 97.40** |
| A25 | 0.75* | 0.27 NS | 98.97 NS |
| B65 | 0.07 NS | 0.64* | 99.30* |

*Statistical significance at $P < 0.05$; **statistical significance at $P < 0.01$; NS, nonsignificant

[a] Population clusters: Europe (Albania, Andalusia, Basques, Basque Guipúzcoa, Basque Navarre, Catalonia, Canary Islands, France, Greece, Macedonia, Romania, Switzerland, and Valencia); Africa (Algeria, Sahara, Tunisia, North-, Southeast-, and West-Morocco); Caucasus (Armenia, Azerbaijan, Cherkesia, Georgia and Kabardinia). Sources: Stoneking et al. (1997), Romualdi et al. (2002) and Comas et al. (2000, 2004)

[b] $F_{CT}$, genetic variation among groups

[c] $F_{SC}$ genetic variation among populations within groups

[d] $F_{ST}$, genetic variation among individuals within populations
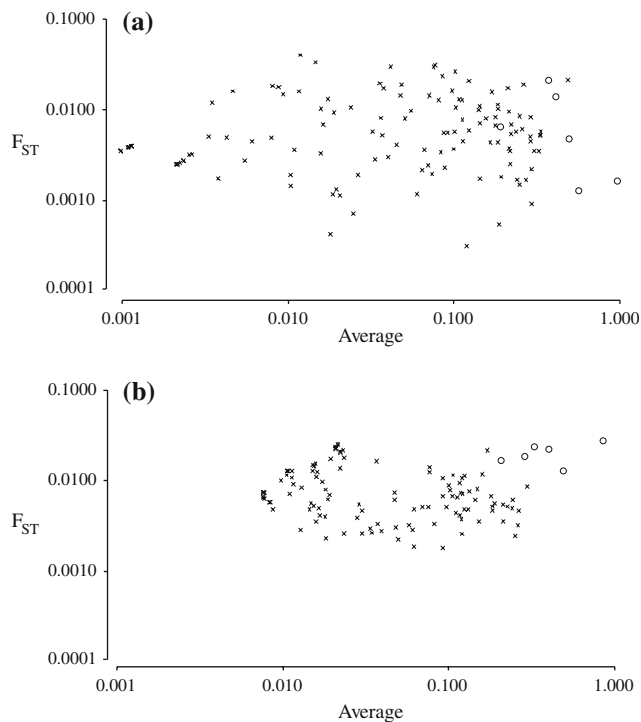
**Fig. 3** Scatterplot of average frequencies vs. Wright's $F_{ST}$ values for autosomal microsatellite ($x$) and *Alu* ($o$) alleles. **a** Considering only samples from different provinces of the Iberian Basque area, **b** including other European samples, in addition of Basque subpopulations

## Discussion

In human evolutionary studies, paternal (Y-chromosome) and maternal (mtDNA) lineages have been extensively analyzed to characterize human groups in terms of origin, demographic milestones with genetic implications (founder effects, bottlenecks) and genetic variability. However, the reconstruction of the history of human populations from genetic data is a complex task that also requires information from the recombining parts of the nuclear DNA, namely, the autosomes (Kidd et al. 2000). Polymorphic *Alu* insertions represent an important source of nuclear genetic diversity. *Alu* elements are highly stable markers that are not affected by substantial mutation rates. Consequently, *Alu* repeats form a rich molecular fossil record that is faithfully recorded in the human genome from generation to generation, thereby representing an advantageous characteristic in phylogenetic studies. In addition, *Alu* elements are widely dispersed throughout the human genome, subject to extremely limited amounts of gene conversion, and selectively neutral (Batzer et al. 1996; Comas et al. 2000). In terms of selection, Cordaux et al. (2006) have suggested that most young *Alu* elements can be considered to be neutral residents of the human genome. With mutation and selection ruled out, the analysis of *Alu* insertions should solely reflect processes of interaction between gene flow and genetic drift in the past.

The genetic structure emerging from MDS and AMOVA analyses indicates a substantial heterogeneity among the populations considered, which is basically distributed between three different geographic areas: the Caucasus, the rest of Europe and North Africa. It is worth underscoring the presence of a significant *Alu* diversity among the Caucasus populations. The high degree of genetic heterogeneity observed in the Caucasus region is probably caused by effect of genetic drift, which is promoted by the small population sizes and the relative isolation resulting from the complex orography of this geographic zone, as has been suggested in previous works (Nasidze et al. 2001; Alfonso-Sánchez et al. 2006). This reasoning is also a valid explanation for the differentiation between Caucasus and European populations.

In the MDS plot, the samples representing the Basque Country appear to be clearly distanced from the Caucasus populations. This result conflicts with previous findings regarding the variability of the immunoglobulin (GM and KM) genes, whose authors enunciated a hypothesis linking the origin of Basques with a small Neolithic North Caucasian population (Calderón et al. 1998). Within this global picture of *Alu* heterogeneity between Basque samples and populations from the Caucasus region, particular interest should be given to the scanty genetic affinity between the Basques and the Georgians. Both of these human groups are considered to be linguistic isolates. It has been suggested that Euskera (Basque language) and Kartvelian (Swanetian language) are the remnants of pre-Indo-European languages of Paleolithic antiquity (Renfrew 1991). A recent linguistic classification has the Euskera sharing the same cluster with the Caucasian languages (Chen et al. 1995). Based mainly on such linguistic criteria, some experts have often postulated that Basques and Georgians have a common Upper Paleolithic background (Gamkrelidze and Ivanov 1990; Ruhlen 1991). However, *Alu* markers examined in the present study failed to detect close genetic affinity between these groups, as can be inferred from their positions in the two-dimensional representation. It can be concluded, therefore, that *Alu* data do not support the putative relationship between language and genes in Basques and Georgians. This finding is compatible with postulates of several previous studies, based either on classical markers (Bertorelle et al. 1995) or on DNA molecular markers (Alfonso-Sánchez et al. 2006). Likewise, the genetic

information derived from the analysis of *Alu* elements does not sustain the notion of a common origin between Basques and North Africans (Berbers), as has been proposed in earlier population genetic surveys (see Arnaiz-Villena et al. 1997). In fact, the samples with Berber origins (North Morocco and Southeastern Morocco) appear to be interspersed within the North African cluster and clearly segregated from the Basques.

The genetic heterogeneity found between North Africa and Europe based on *Alu* diversity is consistent with results reported in previous studies using classical genetic markers (Bosch et al. 1997; Simoni et al. 1999), microsatellites or STRs (Bosch et al. 2000), Y-chromosome STRs (Bosch et al. 1999; Manni et al. 2002; Semino et al. 2004), polymorphic *Alu* insertions (Comas et al. 2000, García-Obregón et al. 2006) and HLA-class II loci (Pérez-Miranda et al. 2003, 2004). Several investigations claim that the genetic discontinuity among populations from the two Mediterranean shores could be related to a strong barrier to gene flow between Africa and Europe at the Strait of Gibraltar (Simoni et al. 1999; Comas et al. 2000; Manni et al. 2002). It has also been argued that a Mesolithic (or older) in situ differentiation of the human groups established in northwestern Africa accounts for the abovementioned genetic differentiation (Bosch et al. 1997).

With respect to the group of European populations (excluding the Caucasus region), no clear geographic structuring of the *Alu* diversity was observed; therefore, the notion of heterogeneity due to isolation-by-distance can be ruled out. The Basque collections included in the MDS analysis remain grouped with the vast majority of European populations. Both the MDS and the AMOVA analyses based on *Alu* data denote a lack of significant genetic heterogeneity, both among Basque subpopulations and among autochthonous Basques and other neighboring populations. However, there is a sizeable number of publications which report that the Basques have gene frequencies that are notably different from the expected ones for a population theoretically inserted—according to geographic location—within the European genetic landscape. In such studies, a wide range of genetic markers have been analyzed, including blood group systems *ABO* (Boyd and Boyd 1937), *Rh* (Etcheverry 1945) and *Duffy* (Levine et al. 1977), some Y-chromosome DNA haplotypes (Lucotte and Hazout 1996), mtDNA haplogroups (Torroni et al. 1998, 2001), immunoglobulin allotypes (Calderón et al. 1998), HLA class-II genes (Pérez-Miranda et al. 2003, 2004) and autosomal STRs (Pérez-Miranda et al. 2005), among others. Specifically

within the scope of the Basque area, the bulk of the population genetic studies identify spatial substructuring of the Basque gene pool (Goedde et al. 1972, 1973; Aguirre et al. 1991; Calderón et al. 1998; Pérez-Miranda et al. 2003, 2005, among others), although this topic has been questioned in a few publications (Calafell and Bertranpetit 1994; Comas et al. 1998).

In an attempt to analyze the discrepancies in the degree of genetic heterogeneity detected by different types of polymorphism in the Basque gene pool, Wright's $F_{ST}$ values computed from both *Alu* insertions and from other genetic loci not submitted to selective pressures (autosomal STRs) were compared in two different databases: (1) only subpopulations of the Iberian Basque Country, (2) other European samples in addition of the Basque subpopulations. At the continental scale, *Alu* insertions showed a genetic heterogeneity visibly higher than the obtained one for autosomal STRs, whereas in the context of the Basque area the values of Wahlund's variance were similar for both types of polymorphic markers.

When we consider a broad geographic region inhabited by a group of populations with a relatively ancient common ancestor, *Alu* insertions are expected to have higher values of Wright's $F_{ST}$ than the autosomal STRs. The microsatellites are characterized by relatively high mutation rates that oscillate between $3.3 \times 10^{-4}$ per locus per generation (25 years in a generation) (Forster et al. 2000) and $15.2 \times 10^{-4}$ per locus per generation (Zhivotovsky et al. 2004). The combined action of the high mutation rate of human microsatellites with the homogenizing effect typical of gene flow may eliminate or dilute the genetic information on the origins of a given population by blurring any signal of phylogenetic affinity with ancient human groups. On the contrary, because *Alu* insertions are unique events and, consequently, much more conservative, they are not exposed to the random fluctuations caused by mutation action. Therefore, *Alu* elements will better reflect the common ancestral origin of the populations of a given geographic region. Within the context of the Basque area, polymorphic *Alu* insertions showed the same level of genetic heterogeneity as the autosomal STRs. For this reason, we suggest that the spatial substructuring detected in the Basque gene pool from the analysis of diverse genetic markers (see references above) could have had a relatively recent origin. In effect, if the population subdivision unveiled by a group of polymorphic markers had begun to take place at some stage near to the origin of the Basque population, this fact should be reflected in Wright's $F_{ST}$ have greater values for *Alu* insertions than for microsatellites. The observable differences for some gene

frequencies between the provinces of the Basque Country probably stem from recent admixture processes with surrounding populations; these processes would occur in variable degrees for each Basque region according to the predominance of the Basque language.

The Basque population is a genetic outlier in Europe. As we have seen above, many studies have exposed the singularities of the genetic background of the autochthonous Basque population based on both classical genetic markers and DNA molecular markers (reviews in de Pancorbo et al. 2001; Pérez-Miranda et al. 2005). In explaining the genetic distinctiveness of Basques, the most common scenario is random genetic drift and high inbreeding levels over long periods in association with an isolation from surrounding populations. Because Basques represent a linguistic island within the Iberian Peninsula, some authors have suggested that the genetic isolation of the Basques may be a consequence of their peculiar language (Cavalli-Sforza et al. 1994; Calderón et al. 1998; de Pancorbo et al. 2001). Language is considered one of the major sociocultural factors restricting gene flow and population admixture in that it prevents the assimilation of immigrants into the native (recipient) population and increases ethnic endogamy, whose main consequence would be the departure of panmixia (Alfonso-Sánchez et al. 2001, 2005). Indeed, it is well documented that linguistic boundaries can be effective barriers to gene flow (Barbujani and Sokal 1990; Barbujani 1997). This seems to be the reason why Guipúzcoa province, which does not border on other non-Basque (Romance-speaking) regions, often appears as the most differentiated Basque subpopulation in comparison to other Iberian and even Basque populations (see Fig. 1), as has been confirmed in different previous works (Calderón et al. 1998; de Pancorbo et al. 2001; Pérez-Miranda et al. 2005).

In summary, our analysis of eight *Alu* insertions revealed no significant genetic heterogeneity between Basque subpopulations or between Basques and other European populations. Based on a comparison of the genetic structure of the Basque subpopulations obtained from *Alu* insertions and autosomal STRs, it can be inferred that distinct polymorphic markers will reveal distinct domains of the evolutionary history of human populations, depending on the particular characteristics of each genetic loci in terms of level of polymorphism, mutation rate, potential action of back mutation, etc. The genetic heterogeneity between Basque subpopulations identified in several previous polymorphism studies seems to be relatively recent and caused by a differential process of genetic admixture

with non-Basque neighboring populations, modulated by the effect of a linguistic barrier to random mating (panmixia). Although it cannot be stated with absolute certainty that the origin of the Basques is recent and that its gene pool is modeled solely by interplay between the isolation resulting from the linguistic barrier to panmictic matings and the effect of genetic drift, our analyses do support this hypothesis as the most plausible, bearing in mind the manifest contradiction of the *Alu* data presented herein and the hypotheses linking the origin of Basques with the North of Africa or the Caucasus region. It would be interesting to extend the number of *Alu* insertions studied in order to complement the genetic information provided by analyses of maternal (mtDNA) and paternal (Y-chromosome) lineages and reveal new clues about the origin and past demographic interactions of the Basque population.

## References

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

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

Alfonso-Sánchez MA, Martínez-Bouzas C, Castro A, Peña JA, Fernández-Fernández I, Herrera RJ, de Pancorbo MM (2006) Sequence polymorphisms of the mtDNA control region in a human isolate: the Georgians from Swanetia. J Hum Genet 51:429–439

Arnaiz-Villena A, Martinez-Laso J, Gomez-Casado E, Diaz-Campos N, Santos P, Martinho A, Breda-Coimbra H (1997) Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes. Immunogenetics 47:37–43

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

Batzer MA, Deininger PL (2002) *Alu* repeats and human genomic diversity. Nature 3:370–379

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD,

Herrera RJ, Deininger PL (1994) African origin of human specific polymorphic *Alu* insertions. Proc Natl Acad Sci USA 91:12288–12292

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M (1996) Genetic variation of recent *Alu* insertions in human populations. J Mol Evol 42:22–29

Bertorelle G, Bertranpetit J, Calafell F, Nasidze IS, Barbujani G (1995) Do Basque- and Caucasian-speaking populations share non-Indo-European ancestors? Eur J Hum Genet 3:256–263

Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, Bertranpetit J (1997) A population history of Northern Africa: Evidence from classical genetic markers. Hum Biol 69:295–311

Bosch E, Calafell F, Santos FR, Pérez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623–1638

Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, Mateu E, Martínez-Arias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Hariti G, Cambon-Thomsen A, Bertranpetit J (2000) Genetic structure of North-West Africa revealed by STR analysis. Eur J Hum Genet 8:360–366

Boyd WC, Boyd LG (1937) New data on blood groups and other inherited factors in Europe and Egypt. Am J Phys Anthropol 23:49–70

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 A, 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, Princeton

Chen J, Sokal RR, Ruhlen M (1995) Worldwide analysis of genetic and linguistic relationships of human populations. Hum Biol 67:595–612

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

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer M Bertranpetit J, Sajantila A (2000) 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, Plaza S, Calafell F, Sajantila A, Bertranpetit J (2001) Recent insertion of an Alu element within a polymorphic human-specific *Alu* insertion. Mol Biol Evol 18:85–88

Comas D, Schmid H, Braeuer S, Flaiz C, Busquets A, Calafell F, Bertranpetit J, Scheil HG, Huckenbeck W, Efremovska L, Schmidt H (2004) *Alu* insertion polymorphisms in the Balkans and the origins of the Aromuns. Ann Hum Genet 68:120–127

Cordaux R, Jungnam L, Dinoso L, Batzer MA (2006) Recently integrated *Alu* retrotransposons are essentially neutral residents of the human genome. Gene 373:138–144

Etcheverry MA (1945) El factor rhesus en genetica e importancia clinica. Diagn Med 17:1237–1251

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinf Online 1:47–50

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67:182–196

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

García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Vidales C, Arroyo D, Peña JA (2006) Genetic position of Valencia (Spain) in the Mediterranean basin according to *Alu* insertions. Am J Hum Biol 18:187–195

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

Jorde L, Bamshad M, Watkins W, Zenger R, Fraley A, Krakowiak P, Carpenter K, Soodyall H, Jenkins T, Rogers A (1995) Origins and affinities of modern humans: A comparison of mitochondrial and nuclear data. Am J Hum Genet 57:523–538

Jorde L, Rogers A, Bamshad M, Watkins W, Krakowiak P, Sung S, Kere J, Harpending H (1997) Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103

Kidd KK, Kidd JR, Pakstis AJ, Bonné-Tamir B, Grigorenko E (2000) Nuclear genetic variation of European populations in a global context. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. The McDonald Institute for Archaeological Research, Cambridge, pp 109–117

Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:115–129

Lalouel JM (1973) Topology of population structure. In: NE Morton (ed) Genetic structure of populations. University of Hawaii Press, Honolulu, pp 139–149

Lell JT, Wallace DC (2000) The peopling of Europe from the maternal and paternal perspectives. Am J Hum Genet 67:1376–1381

Levine MH, Von Hagen V, Ruffié J, Darrasse H (1977) A hematological approach to basque isolation in two french Basque villages. Ann NY Acad Sci 293:185–193

Liu K, Muse S (2004) POWERMARKER: New genetic data analysis software. Version 3.0. Free program distributed by the author over the internet from http://www.powermarker.net

Lucotte G, Hazout S (1996) Y-chromosome DNA haplotypes in Basques. J Mol Evol 42:472–475

Maca-Meyer N, Villar J, Pérez-Méndez L, Cabrera de León A, Flores C (2004) A tale of aborigines, conquerors and slaves: Alu insertion polymorphisms and the peopling of Canary Islands. Ann Hum Genet 68:600–605

Manni F, Leonardi P, Barakat A, Rouba H, Heyer E, Klintschar M, McElreavey K, Quintana-Murci L (2002) Y-chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Africa. Hum Biol 74:645–658

Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M (2001) *Alu* insertion polymorphisms and the genetic structure of human populations from the Caucasus. Eur J Hum Genet 9:267–272

de Pancorbo MM, López-Martínez M, Martinez-Bouzas C, Castro A, Fernández-Fernández I, Antúnez de Mayolo G,

Antúnez de Manolo A, Antúnez de Manolo P, Rowold DJ, Herrera RJ (2001) The Basques according to polymorphic *Alu* insertions. Hum Genet 109:224–233

Peña JA, Morales B, Calderón R (1997) New method for comparing levels of microdifferentiation: application to migration matrices of two populations from the Basque Country (Spain). Hum Biol 69:329–344

Peña JA, Calderón R, Pérez-Miranda A, 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

Peña JA, García-Obregón S, Pérez-Miranda AM, De Pancorbo MM, Alfonso-Sánchez MA (2006) Gene flow in the Iberian Peninsula determined from Y-chromosome STR loci. Am J Hum Biol 18:532–539

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, Vidales C, Calderón R, Peña JA (2004) Genetic polymorphism and linkage disequilibrium of the HLA-DP region in Basques from Navarre (Spain). Tissue Antigens 64:264–275

Pérez-Miranda AM, Alfonso-Sánchez MA, Kalantar A, García-Obregón S, de Pancorbo MM, Peña JA, Herrera RJ (2005) Microsatellite data support subpopulation structuring among Basques. J Hum Genet 50:403–414

Renfrew C (1991) Before Babel, speculations on the origins of linguistic diversity. Cambridge Archaeol J 1:3–23

Reynolds J, Weir BS, Cockerman CC (1983) Estimation of the coancestry coefficient: bases for a short term genetic distance. Genetics 105:767–779

Richards M (2003) The neolithic invasion of Europe. Annu Rev Anthropol 32:135–162

Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 12:602–612

Ruhlen M (1991) A guide to the world's languages (2nd edn). Stanford University Press, Stanford

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034

Simoni L, Gueresi P, Pettener D, Barbujani G (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. Hum Biol 71:399–415

Sokal RR, Rohlf FJ (1997) Biometry. The principles and practice of statistics in biological research, 3rd edn. WH Freeman, New York

Stoneking M, Fontius J, Clifford S, Soodyall H, Arcot S, Saha N, Jenkins T, Tahir M, Deininger P, Batzer MA (1997) *Alu* insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. Genome Res 7:1061–1071

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonné-Tamir B, Scozzari R (1998) MtDNA analysis reveals a major late paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137–1152

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R (2001) A signal from human mtDNA of postglacial recolonization in Europe. Am J Hum Genet 69:844–852

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB (2001) Patterns of ancestral human diversity: an analysis of *Alu*-insertion and restriction-site polymorphisms. Am J Hum Genet 68:738–752

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AE, Carroll ML, Nguyen SV, Walker JA, Prasad BVR, Reddy PG, Das PK, Batzer MA, Jorde LB (2003) Genetic variation among world populations: Inferences from 100 *Alu* insertion polymorphisms. Genome Res 13:1607–1618

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