

Microsatellites and *Alu* elements from the human MHC in Valencia (Spain): analysis of genetic relationships and linkage disequilibrium

S. García-Obregón*, M. A. Alfonso-Sánchez*, L. Gómez-Pérez*, A. M. Pérez-Miranda*, D. Arroyo†, M. M. de Pancorbo‡ & J. A. Peña,*

Summary

Two different sets of noncoding markers (microsatellites and *Alu* elements) from the human chromosome six were analysed in 106 individuals from Valencia (Spain), with the aim of exploring the effect of evolutionary forces on the genetic variability of the major histocompatibility complex (MHC) and assessing the potential usefulness of these genetic loci in phylogenetic studies. Linkage disequilibrium (LD) analyses revealed statistically significant associations among markers located in the MHC class I region, and also between the microsatellite D6S2792 and several genetic loci from MHC class I, II and III regions. Results of the Ewens-Watterson test indicated that only D6S2792 showed significant departure from selective neutrality. Despite the paucity of haplotype data in the literature, results of the phylogenetic analyses at world scale (*Alu* elements) showed that the genetic relationships of Valencia were mainly determined by the ethnic ancestry of the populations considered, whereas at European scale (microsatellites) population affinities were strongly influenced by geography. Our findings suggest that noncoding markers from the MHC such as *Alu* and microsatellite loci might have a potential value as lineage (ancestry) markers in investigations into evolutionary, medical and forensic perspectives.

Introduction

The gene family encoding proteins involved in the human major histocompatibility complex (MHC) has been mapped to the short arm of the chromosome 6

(6p21.3). It contains more than 150 expressed genes that are clustered in three major genomic regions, namely class I, II and III (Shiina *et al.*, 2009). Human leucocyte antigen (HLA) genes are included these genomic regions. Of these MHC genes, about 40% are postulated to be involved in immune regulation (MHC sequencing consortium, 1999; Loisel *et al.*, 2008). Numerous authors have suggested that the physiological role of the HLA genes could be one of the main causes contributing to the maintenance of the high levels of polymorphism of the MHC region (Meyer & Thomson, 2001; McClland *et al.*, 2003; Segal & Hill, 2003; Piertney & Oliver, 2006).

In addition to HLA genes, other types of DNA sequences coexist in the MHC: transposable elements (transposons, retrotransposons), regulatory elements, pseudogenes and other noncoding regions such as minisatellites, microsatellites, single nucleotide polymorphisms (SNPs) and the *Alu* family of repetitive elements (Shiina *et al.*, 2009). Polymorphic *Alu* insertions are among the more valuable noncoding markers in molecular studies with evolutionary perspective. They are DNA sequences of approximately 300 base pairs (bp) ancestrally originated by retrotransposition of the 7SL RNA gene, with which *Alu* elements share about 90% of sequence homology. *Alu* insertions are the most common short interspersed elements (SINEs): they represent close to 11% of the human genome, with around one million copies per haploid genome (Batzer & Deininger, 2002). Human-specific *Alu* members possess some exceptional characteristics to investigate human evolution and conduct population genetic studies. These properties are, for instance, the existence of a single common mutation for all chromosomes with the same insertion (identity by descent), lack of insertion as the ancestral state, and no known mechanism for the complete and specific removal of an *Alu* element. (Batzer & Deininger, 1991; Batzer *et al.*, 1994; International Human Genome Sequencing Consortium, 2001).

Because of their hypervariable nature, microsatellites are one of the most informative noncoding markers for genetic mapping, linkage analysis, human identity testing and for the elucidation of human demographic

* Departamento de Genética y Antropología Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco, Bilbao, Spain, † Progenie Molecular, Valencia, Spain, and ‡ BIOMICs Research Group, Centro de Investigación y Estudios Avanzados "Lucio Lasca-ray", Universidad del País Vasco, Vitoria-Gasteiz, Spain

Received 9 February 2011; revised 17 June 2011; accepted 11 August 2011

Correspondence: José A. Peña, Departamento de Genética, Antropología Física y Fisiología Animal, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain. Tel: +34 94601 26 00; Fax: +34 94601 35 00; E-mail: joseangel.pena@ehu.es

evolution (Rowold & Herrera, 2005). Microsatellites are DNA sequences with short specific motifs repeated numerous times in a head to tail manner (Hearne *et al.*, 1992). They are scattered all over the genome and constitute highly polymorphic loci with numerous codominant alleles differing in the number of copies of the repeated motif (Beckman & Weber, 1992; Tautz & Schlötterer, 1994). At present, there are about 20 000 known hypervariable microsatellite loci in the human genome. Specifically, in the human MHC, a total of 1527 microsatellites have been identified, of which at least 268 have been utilized in disease association studies (Shiina *et al.*, 2009). Microsatellites have become one of the most powerful tools in human population genetics, owing to characteristics such as ease in genotyping and scoring, large number of alleles, high heterozygosity and widespread distribution within the human genome.

The high genetic variability of the human MHC has been widely employed in the search of DNA molecular markers associated with diseases, especially those of autoimmune origin (Lie & Thorsby, 2005; Shiina *et al.*, 2009) such as rheumatoid arthritis (Thompson *et al.*, 2004). For the same reason, this genomic region has been the focus of a number of phylogenetic and evolutionary studies in human populations (Sanchez-Mazas, 2001; Pérez-Miranda *et al.*, 2003, 2004).

In this work, polymorphism of five *Alu* loci (*AluMicB*, *AluTF*, *AluHJ*, *AluHG* and *AluHF*) and five microsatellite markers (D6S291, D6S2666, D6S2792, D6S265 and D6S105) from the MHC region (see Fig. 1) was screened in a Mediterranean sample from Valencia province (Spain). Findings of previous studies have revealed strong associations between MHC *Alu* markers and HLA genes (Dunn *et al.*, 2005, 2007; Kulski & Dunn, 2005; Tian *et al.*, 2008; Yao *et al.*, 2009). Likewise, other investigations found that the phylogenetic information provided by HLA genes is quite well reproduced by adjacent noncoding markers such as microsatellites (García-Obregón *et al.*, 2010). Thus, our study was intended to characterize the polymorphism of the targeted *Alu* and microsatellite markers to gain insights about the effect of evolutionary forces on the genetic variability of neutral noncoding

markers from the MHC region. In addition, population databases at different geographic scales (according to data availability) were constructed for both types of molecular markers to examine the genetic relationships of the Valencian sample in a broader geographic context and to evaluate the potential usefulness of the genetic information generated by MHC noncoding markers in phylogenetic studies.

Materials and methods

Sample

The Autonomous Community of Valencia is a political/administrative entity located in the Spanish Mediterranean coastal region, in eastern Iberia. It comprises the provinces of Alicante, Castellón and Valencia, stretching over 474 km of coastline and occupying a total area of 23 305 km². Current population density of Valencia is more than twice the average for Spain at 178.6 inhabitants/km².

DNA samples were obtained from 106 unrelated, healthy individuals from Valencia province. Birthplace of the voluntary donors, surveyed back to the third generation (the grandparents), was the criterion to decide its inclusion in the sample. All volunteers gave their informed consent prior to inclusion in the sample, following the ethical guidelines stipulated for research with human beings. The study protocol was approved by the Institutional Review Board from Universidad del País Vasco.

Alu typing

Genomic DNA was extracted from bloodstains using a Qiagen kit (QIAmp DNA Micro Kit, Qiagen, Hilden, Germany) and stored at -20°C. Five autosomal *Alu* insertions (*AluMicB*, *AluTF*, *AluHJ*, *AluHG* and *AluHF*) were typed in the sample population. PCRs were performed in a final volume of 10 µL using a 20 ng of DNA, PCR buffer (50 mM of KCl, 10 mM of Tris-HCl and 0.01% gelatine), 3 mM of MgCl₂, 1 mM of each dNTP, 2.5 µM of each primer and 0.5 U Taq polymerase in a Thermal Cycler Gene Amp PCR

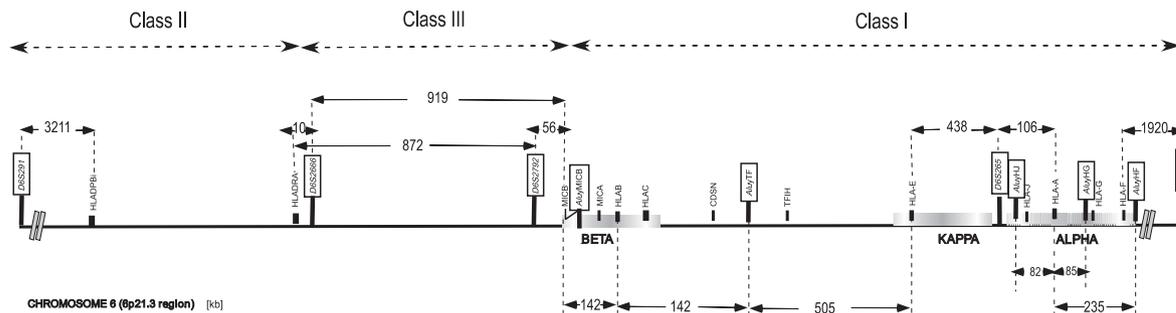


Figure 1. Chromosome 6 (6p21.3 region) map showing the position of 10 microsatellite and *Alu* markers located at the major histocompatibility complex region. Figures represent distance between markers and human leucocyte antigen (HLA) or HLA-like genes in kilobases.

System 9700 (Applied Biosystems, Foster City, CA, USA), as previously described by (Kulski *et al.*, 2001, 2002; Dunn *et al.*, 2002, 2003, 2007; Kulski & Dunn, 2005). Each sample was subjected to the following amplification conditions: 35 subsequent cycles consisted of denaturing at 94°C for 30 s followed by annealing at different temperatures depending on the marker 45 s and an extension of 45 s at 72°C. NCBI accession numbers, primer sequences, annealing temperatures and references for each *Alu* are presented in Table S1 (Supporting Information). All PCRs were directly electrophoresed in 1.5% agarose gels stained with ethidium bromide (0.5 µL/ml), viewed under UV light and documented using digital photography. To control the quality of the PCR and electrophoresis, we included both a positive control (homozygote for the insertion) and a negative control containing all PCR reagents except template DNA. In addition, a DNA ladder was used to determine the size of the alleles in the electrophoresis gel.

Microsatellite typing

For each sample, a set of five microsatellite loci (D6S291, D6S2666, D6S2792, D6S265 and D6S105) was typed. PCRs were performed in a final volume of 10 µl using a 50–100 ng of DNA, PCR buffer (50 mM of KCl, 1.5 mM of MgCl₂, 10 mM of Tris-HCl and 0.01% gelatine), 1.5 mM of MgCl₂, 0.2 mM of each dNTP, 0.5 µM of each primer and 0.5 U Taq polymerase in a Thermal Cycler Gene Amp PCR System 9700 (Applied Biosystems) (García-Obregón *et al.*, 2010). PCR amplification conditions and primer sequences utilized for each microsatellite are given in Table S2 (Supporting Information). Amplified fragments were detected and separated by capillary electrophoresis, using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). An internal size standard (GeneScan500 TAMRA; Applied Biosystems) was included. Fragment sizes were determined automatically using Genescan Analysis software version 3.1 and by comparison with the supplied allelic ladders. Positive and negative PCR controls were performed to test the quality of the amplification process.

Statistical analysis

Genotype and allele frequencies for all markers were calculated by direct counting. To assess the genetic variability of the microsatellites and *Alu* elements examined, expected heterozygosity (H_e) and polymorphic information content (PIC) were computed using the Power Marker v. 3.25 program (Liu & Muse, 2005). To test for Hardy–Weinberg equilibrium (HWE) expectations, a Fisher's exact probability test was conducted to estimate P-values (Guo & Thompson, 1992) using the Arlequin v. 3.5 program (Excoffier & Lischer, 2010). Bonferroni correction for multiple comparisons was applied.

Linkage disequilibrium analysis and haplotype frequencies

Given that both types of molecular markers are located in chromosome 6, linkage disequilibrium values were computed for each loci pair using Arlequin v. 3.5 program (Excoffier & Lischer, 2010). Haplotypes were constructed using genotype data for each individual based on five *Alu* elements (*Alu*MicB-*Alu*TF-*Alu*HJ-*Alu*HG-*Alu*HF). Additionally, haplotypes based on 10 loci (five microsatellites plus five *Alu* elements) were estimated. All haplotypes were inferred, and their frequencies were calculated with the ARLEQUIN v3.5 software.

Selection studies

To evaluate the impact of potential selective pressures influencing on microsatellite and *Alu* frequencies, departure from selective neutrality was assessed using the Ewens–Watterson test (Ewens, 1972; Watterson, 1978).

Phylogenetic studies

To analyse genetic affinities of the Valencian sample in a broader geographic context, five *Alu* loci haplotype frequency data for 11 worldwide populations were compiled from the available relevant literature (Kulski & Dunn, 2005; Tian *et al.*, 2008) (see Table S3, Supporting Information). *Alu* haplotype frequencies were then used to compute F_{ST} genetic distances (Reynolds *et al.*, 1983) between all pairs of populations with the PHYLIP v3.2 program (Felsenstein, 2005). Nonmetric multidimensional scaling (MDS) analysis was further performed using the SPSS v13.5 statistical package (SPSS Inc., Chicago, IL, USA) to represent the resultant F_{ST} genetic distance matrix in a two-dimensional space.

Because of the remarkable scarceness of MHC microsatellite population databases, microsatellites D6S105, D6S291 and D6S2666 were excluded from the phylogenetic analysis. For the same reason, this analysis was restricted to the European context. We followed the same procedure as in *Alu* insertions: construction of a database for 15 European populations (including Valencia) by compiling allelic frequency data from previous publications (Charron, 1997; Peña *et al.*, 2002; García-Obregón *et al.*, 2010) (see Table S4, Supporting Information), calculation of F_{ST} genetic distances (Reynolds *et al.*, 1983) between pairs of populations based on allele frequencies for D6S2792 and D6S265, and representation of the F_{ST} genetic distance matrix in a two-dimensional space through MDS.

Results

Genotype and allele frequencies for the five *Alu* loci typed in Valencia are listed in Table 1. All *Alu* loci

were polymorphic, i.e. neither fixed insertions nor insertions absent from the sample examined were found. Concerning microsatellites (Table 2), we identified a total of seven alleles for the D6S2666 locus, eight alleles for D6S291, nine alleles for D6S265 and 12 each for the D6S105 and D6S2792 microsatellites. The most frequent alleles for microsatellite markers were 121 bp for D6S105, 127 bp for D6S265, 99 bp for D6S2792, 170 bp for D6S291 and 170 bp for D6S2666. Genotype frequencies are given in Table S5.

The degree of genetic variability of the molecular markers considered was assessed by computing expected heterozygosity (He) and polymorphic information content (PIC) for each locus. As expected, the highest values for both diversity parameters were obtained for microsatellites (Table 3), notably D6S2792 (He: 0.875, PIC: 0.862). Likewise, D6S2666 proved to be the least informative marker among microsatellites, showing the lowest He and PIC values. As regards to biallelic *Alu* markers, He values oscillated between 0.286 (*Alu*Mic B) and 0.393 (*Alu*HF), whereas PIC values did not exceed 0.32 in any case. After applying the conservative Bonferroni correction for multiple testing, no significant departure from HWE expectations was detected in any of the *Alu* and microsatellite markers.

Linkage disequilibrium and selection test

Results generated by the linkage disequilibrium (LD) analysis for microsatellites and *Alu* elements are schematically represented in Fig. 2. A substantial part of the significant LD values was obtained between loci located in the MHC class I region (*Alu*TF, *Alu*HJ, D6S265, *Alu*HG and *Alu*HF), whereas the remaining significant LD values involved different MHC class I, II and III markers (D6S2792 and D6S291, D6S2666, *Alu*MicB, *Alu*TF and D6S105). Departure from selective neutrality was tested in both *Alu* and microsatellite markers by considering in each case those populations included in the phylogenetic analysis. D6S2792 microsatellite was the only marker that featured a significant departure from selective neutrality, according to the results provided by the Ewens–Watterson test ($P = 0.008$).

Haplotype frequency distribution

In this study, five *Alu* loci and ten loci (microsatellites plus *Alu* markers) haplotypes were constructed. Haplotype's definition and frequencies for the combination of the five *Alu* insertions are presented in Table 4. Under assumption of random assortment of alleles, 32 different haplotypes are expected; however, only 19 haplotypes were identified in the sample examined. The most frequent haplotype proved to be the lack of insertion in all *Alu* loci, with a frequency value above 28%. Likewise, the haplotype formed by the insertion of all *Alu* elements was not observed in Valencia collection. To the best of our knowledge, this haplotype has not been identified in the population analysed so far, which seems to situate the origin of these insertions in a relatively recent evolutionary period.

The haplotype frequencies considering ten loci (*Alu* + microsatellites) can be consulted in Table S6 (Supporting Information).

Phylogenetic analysis

Bearing in mind the paucity of haplotype frequency data for the MHC *Alu* insertions examined in this study, phylogenetic analysis was performed by considering all databases available for worldwide populations in the relevant literature, which allowed assessing the genetic position of Valencia in a global context. The two-dimensional representation of the F_{ST} genetic distance matrix accounted for 99.3% of the total variance (see Fig. 3), with a relatively low coefficient of stress (4.5%). Genetic affinities among populations strongly mirrored their geographic and/or ethnic characteristics. As expected, Valencia sample plotted close to the Australians, the other population with European ancestry. All the African groups clustered in the negative semiaxis of dimension II, significantly separated from the remaining populations. Asian collections featured certain dispersion all along dimension I, with Thailand relatively close to the European cluster and the centroid of the distribution.

Regarding MHC microsatellites, analysis of population affinities was centred on allele frequencies of D6S2792 and D6S265 microsatellites and circum-

Table 1. Genotype and allele frequencies with standard error (\pm SE) for five *Alu* elements located at chromosome 6 in a population sample from Valencia (Spain)

<i>Alu</i> marker	Genotype frequency			Allele ^a frequency \pm SE		
	1,1	1,2	2,2	1	2	\pm SE
<i>Alu</i> Mic B	0.683	0.288	0.030	0.827	0.173	0.027
<i>Alu</i> TF	0.604	0.337	0.059	0.772	0.228	0.029
<i>Alu</i> HJ	0.575	0.387	0.038	0.769	0.231	0.029
<i>Alu</i> HG	0.650	0.320	0.029	0.811	0.189	0.027
<i>Alu</i> HF	0.548	0.365	0.087	0.731	0.269	0.031

^aAlleles are: 1: lack of *Alu* insertion; 2: presence of *Alu* insertion

Table 2. Allele frequencies and standard error (\pm SE) for five microsatellites located at chromosome 6 in Valencia (Spain)

D6S291		D6S2666		D6S2792		D6S265		D6S105	
allele ^a	Frequency \pm SE								
166	0.015 \pm 0.008	144	0.330 \pm 0.034	97	0.040 \pm 0.013	123	0.140 \pm 0.024	109	0.005 \pm 0.005
168	0.272 \pm 0.030	160	0.038 \pm 0.013	99	0.200 \pm 0.028	125	0.005 \pm 0.005	111	0.005 \pm 0.005
170	0.330 \pm 0.034	164	0.094 \pm 0.020	103	0.080 \pm 0.019	127	0.335 \pm 0.033	113	0.015 \pm 0.008
172	0.068 \pm 0.017	166	0.127 \pm 0.021	105	0.080 \pm 0.018	129	0.125 \pm 0.021	115	0.030 \pm 0.012
174	0.102 \pm 0.022	168	0.019 \pm 0.009	107	0.115 \pm 0.021	131	0.130 \pm 0.023	117	0.205 \pm 0.028
176	0.078 \pm 0.017	170	0.368 \pm 0.032	109	0.120 \pm 0.022	133	0.230 \pm 0.028	119	0.105 \pm 0.022
178	0.126 \pm 0.024	172	0.024 \pm 0.010	111	0.025 \pm 0.011	139	0.010 \pm 0.007	121	0.325 \pm 0.032
182	0.010 \pm 0.007			113	0.005 \pm 0.005	141	0.005 \pm 0.005	123	0.095 \pm 0.020
				115	0.140 \pm 0.023	145	0.020 \pm 0.010	125	0.145 \pm 0.023
				117	0.145 \pm 0.023			127	0.040 \pm 0.013
				119	0.005 \pm 0.005			129	0.025 \pm 0.005
				121	0.045 \pm 0.014			131	0.005 \pm 0.005

^aAllele sizes (in bp).

Table 3. Diversity parameters for microsatellite and *Alu* loci located at the major histocompatibility complex (chromosome 6) in a sample from Valencia (Spain)

Marker	Locus	2N ^a	Observed heterozygosity	Expected heterozygosity	PIC
Microsatellite	D6S291	206	0.748	0.780	0.749
	D6S2666	212	0.717	0.728	0.684
	D6S2792	200	0.920	0.875	0.862
	D6S265	200	0.810	0.782	0.751
	D6S105	200	0.810	0.808	0.784
<i>Alu</i> loci	<i>Alu</i> MicB	202	0.287	0.286	0.245
	<i>Alu</i> TF	202	0.337	0.352	0.290
	<i>Alu</i> HJ	212	0.387	0.335	0.292
	<i>Alu</i> HG	206	0.320	0.307	0.260
	<i>Alu</i> HF	208	0.365	0.393	0.316

PIC, polymorphic information content.

^aSample size in number of chromosomes analysed.

scribed to the European context, also on account of the availability of population databases. MDS plot

accounted for 96.8% of the total variance (Fig. 4), with a coefficient of stress of 9.3%. Results of the MDS mostly overlapped the geographic position of the populations examined. Thus, Valencia plotted close to neighbouring populations from Spain and France (Iberian and French Basques). It is worth mentioning the visibly segregated genetic position of Sardinia, in accordance with several previous studies (Barbujani & Sokal, 1991; Pala *et al.*, 2009), whereas the rest of European populations mostly grouped in the positive semiaxis of dimension II.

Discussion

In this study, two different sets of noncoding markers (polymorphic *Alu* insertions and microsatellites) located in the human MHC genomic region were surveyed in a population sample from Valencia (Spain). As regards the specific case of the MHC *Alu* repeats, our study constitutes the first contribution of frequency data for a human population settled in the European continent.

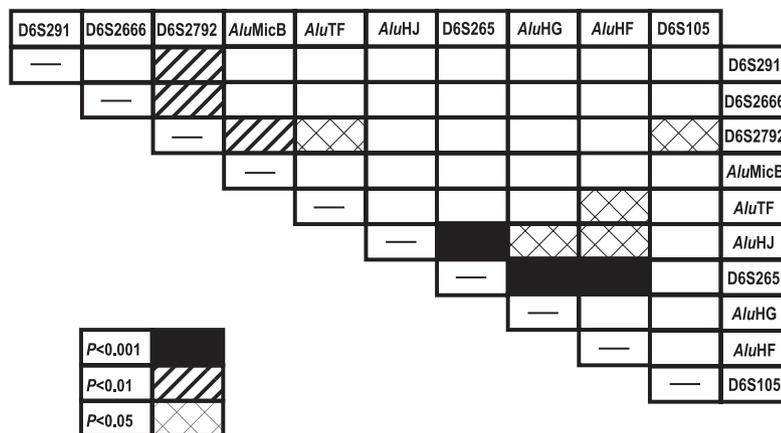


Figure 2. Linkage disequilibrium between all the major histocompatibility complex markers (five microsatellites and five *Alu* insertions) analysed in Valencia population.

Table 4. Haplotype frequencies based on five *Alu* elements (*Alu*MicB, *Alu*TF, *Alu*HJ, *Alu*HG and *Alu*HF) in 106 individuals from Valencia (Spain)

<i>Alu</i> MicB– <i>Alu</i> TF– <i>Alu</i> HJ– <i>Alu</i> HG– <i>Alu</i> HF	Frequency
1-1-1-1-1	0.282
1-1-1-1-2	0.099
1-1-1-2-1	0.068
1-1-1-2-2	0.049
1-1-2-1-1	0.173
1-1-2-1-2	0.010
1-2-1-1-1	0.030
1-2-1-1-2	0.067
1-2-1-2-1	0.042
1-2-2-1-1	0.024
2-1-1-1-1	0.060
2-1-1-2-2	0.016
2-1-2-1-1	0.030
2-2-1-1-1	0.003
2-2-1-1-2	0.023
2-2-1-2-1	0.010
2-2-1-2-2	0.008
2-2-2-1-1	0.001
2-2-2-1-2	0.007

1: lack of *Alu* insertion; 2: presence of *Alu* insertion.

Microsatellites feature relatively high mutation rates that range between 3.3×10^{-4} per locus per generation (Forster *et al.*, 2000) and 15.2×10^{-4} per locus per generation (Zhitovitsky *et al.*, 2004), considering generations of 25 years. The high mutation rate of human microsatellites may thus dilute the genetic information of the origins of a given population by blurring any signal of phylogenetic affinity with ancient human groups. Microsatellites may, in turn, provide valuable genetic information on relatively

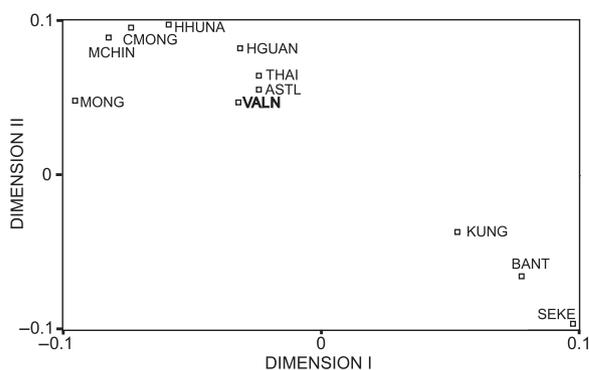


Figure 3. Nonmetric multidimensional scaling applied on *F_{st}* genetic distance matrix. Genetic distances were computed from haplotype frequencies based on combinations of five *Alu* loci (*Alu*MicB, *Alu*TF, *Alu*HJ, *Alu*HG and *Alu*HF), to analyse genetic relationships among 11 worldwide populations. Study population in bold. Population abbreviations: ASTL (Australian Caucasoids), BANT (Bantu), HGUAN (Chinese Guan), HHUNA (Chinese Hunan), CMONG (Chinese Mongolian), KUNG (Kung), MCHIN (Mongolian Chinese), MONG (Mongolian), SEKE (Seke), THAI (Thailand) and VALN (Valencia).

recent demographic events, mainly associated with human migrations, that is, with recent gene flow episodes. On the contrary, *Alu* insertions are presumably unique events (Batzer *et al.*, 1994) and, accordingly, much more conservative, as they are not exposed to the random fluctuations caused by recurrent mutational events. In this way, *Alu* elements will better reflect the common ancestral origin of the populations of a given geographic region. The combined analysis of *Alu* elements and microsatellites may, therefore, provide complementary views of the demographic evolution of human populations.

The absence of geographically comparable databases for both types of markers forced us to carry out population analyses of Valencia at different geographic scopes. Despite this drawback, the geographic scope of each analysis was probably the most appropriate according to the characteristics of the markers examined: a limited geographic area for microsatellites and continental groups for *Alu* insertions. Consequently, results of phylogenetic analysis were statistically robust and congruent with the geographic distribution and/or ethnic ancestry of the populations included. A recent study on the polymorphism of microsatellites from the MHC showed that the genetic information obtained from microsatellites was very similar to that obtained from HLA genes (García-Obregón *et al.*, 2010), so that analysis of the genetic variability of MHC microsatellites could be an appropriate strategy for complementing HLA data in studies on the factors modelling the MHC polymorphism. Regarding *Alu* markers, numerous preceding investigations have demonstrated the usefulness of specific *Alu* repeats as ancestry informative markers, AIMs (Ray *et al.*, 2005; García-Obregón *et al.*, 2007; Terreros *et al.*, 2009; Gómez-Pérez *et al.*, 2010). Human genetic variation tends to be geographically structured, in accordance with historical patterns of gene flow and genetic drift. Our findings reveal the sensitivity of the MHC *Alus* to detect the effects of isolation and genetic drift in the shaping of human genetic variation and, therefore, to differentiate human groups according to ethnic ancestry, as corroborated by the genetic relationships represented in Fig. 3.

Microsatellite D6S2792 was the only one of the noncoding markers examined featuring a significant departure from neutrality, according to the results generated by the Ewens–Watterson test. This microsatellite is located close to the human tumour necrosis factor- α (TNF- α) locus. This gene contains four exons (approximately 3 kb) and encodes a protein (cytokine) with immunomodulatory properties. This is a pleiotropic proinflammatory cytokine that regulates almost all the processes involved in the immune response: activation of monocytes and dendritic cells, adhesion of endothelial cells, organization of lymphoid tissues, etc. (Hajeer & Hutchinson, 2000; Locksley *et al.*, 2001). Bearing in mind this central physiological role, intense selective pressures are expected to

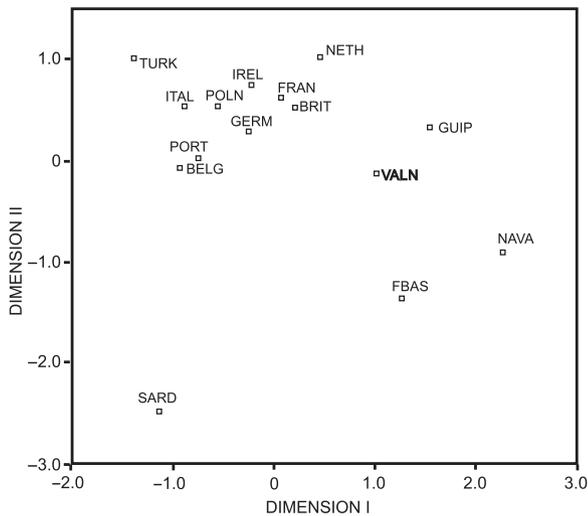


Figure 4. Nonmetric multidimensional scaling applied on F_{ST} genetic distance matrix. Genetic distances were computed from allele frequencies for the major histocompatibility complex microsatellites D6S2792 and D6S265, to analyse genetic relationships among 15 European populations. Study population in bold. Population abbreviations: BELG (Belgium), GUIP (Basques from Guipuzcoa province), NAVA (Basques from Navarre province), BRIT (United Kingdom), FRAN (France), FBAS (French Basques), GERM (Germany), ITAL (Italy), NETH (Netherlands), POLN (Poland), PORT (Portugal), IREL (Ireland), SARD (Sardinia), TURK (Turkey) and VALN (Valencia).

operate upon the TNF- α gene and, therefore, upon closely linked genes from adjacent noncoding DNA regions (as the case of D6S2792) by the effect of the genetic hitchhiking (Hedrick, 1982). Several previous studies have reported a certain association between D6S2792 and susceptibility to some pathologies such as systemic sclerosis (Takeuchi *et al.*, 2000), insulin-dependent (type 1) diabetes mellitus, (Törn *et al.*, 2006), coeliac disease (McManus *et al.*, 1996), psoriasis (Biral *et al.*, 2006) and rheumatoid arthritis (Castro *et al.*, 2001), among others. The existence of selective pressures favouring the maintenance of linked polymorphisms (Mantullo *et al.*, 1995) could also explain the linkage disequilibrium observed between D6S2792 and genetic markers from the MHC class I, II and III regions.

In the quest for understanding population affinities between human groups, as well as ancestry, population genetic profile, demographic history and geographic origin, the construction of biomarker panels based on microsatellites and *Alu* markers represents a reliable, informative and cost-effective experimental method. Therefore, it would be highly advantageous to increase the population databases of MHC *Alus* and microsatellites in future studies, bearing in mind that they can provide interesting complementary information to that obtained from HLA genes for disease association studies, as well as in investigations with evolutionary, epidemiological, medical and forensic perspectives.

Acknowledgements

This work was funded by Research Projects GIU 05/51 from Universidad del País Vasco (UPV/EHU) and IT-424-07 from the Basque Government, Department of Education, Universities and Research. S. García-Obregón was supported through a doctoral fellowship from the Basque Government, Department of Education, Universities and Research.

References

- Barbujani, G. & Sokal, R.R. (1991) Genetic population structure of Italy II. Physical and cultural barriers to gene flow. *American Journal Human Genetics*, **48**, 398.
- Batzer, M.A. & Deininger, P.L. (1991) A human-specific subfamily of *Alu* sequences. *Genomics*, **9**, 481.
- Batzer, M.A. & Deininger, P.L. (2002) *Alu* repeats and human genomic diversity. *Nature*, **3**, 370.
- Batzer, M.A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D.H., Shaikh, T.H. *et al.* (1994) African origin of human specific polymorphic *Alu* insertions. *Proceedings of the National Academy of Science United States of America*, **91**, 12288.
- Beckman, J.S. & Weber, J.L. (1992) Survey of human and rat microsatellites. *Genomic*, **12**, 627.
- Biral, A.C., Magalhaes, F., Wastowski, I.J., Simoes, R., Donadi, E.A., Simoes, A.L., Mendes-Junior, C.T., Tanaka, A.M. & Kraemer, M.H. (2006) Association of HLA-A, -B, -C genes and TNF microsatellite polymorphism with psoriasis vulgaris: a study of genetic risk in Brazilian patients. *European Journal of Dermatology*, **16**, 523.
- Castro, F., Acevedo, E., Ciusani, E., Angulo, J.A., Wollheim, F.A. & Sandberg-Wollheim, M. (2001) Tumour necrosis factor microsatellites and HLA-DRB1*, HLA-DQA1* and HLA-DQB1* alleles in Peruvian patients with rheumatoid arthritis. *Annals of Rheumatic Diseases*, **60**, 791.
- Charron, D. (1997) *Genetic Diversity of HLA. Functional and Medical Implication*. EDK International publisher, Paris.
- Dunn, D.S., Naruse, T., Inoko, H. & Kulski, J.K. (2002) The association between HLA-A alleles and young *Alu* dimorphisms near the HLA-J, -H and -F gene loci in workshop cell lines and Japanese and Australian populations. *Journal of Molecular Evolution*, **55**, 718.
- Dunn, D.S., Inoko, H. & Kulski, J.K. (2003) Characterisation of a dimorphic *Alu* element located between the TFIIF and CDSN genes within the MHC by association studies using workshop cell lines and Japanese and Australian populations. *Electrophoresis*, **24**, 2740.
- Dunn, D.S., Romphruk, A.V., Leelayuwat, C., Bellgard, M. & Kulski, J.K. (2005) Polymorphic *Alu* insertions and their associations with MHC class I alleles and haplotypes in the northeastern Thai. *Annals of Human Genetics*, **69**, 364.
- Dunn, D.S., Choy, M.K., Phipps, M.E. & Kulski, J.K. (2007) The distribution of major histocompatibility complex class I polymorphic *Alu* insertions and their associations with HLA alleles in a Chinese population from Malaysia. *Tissue Antigens*, **70**, 136.
- Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87.
- Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564.

- Felsenstein, J. (2005) *PHYLIP (Phylogeny Inference Package) Version 3.6*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Forster, P., Röhl, A., Lünemann, P., Brinkmann, C., Zerjal, T., Tyler-Smith, C. & Brinkmann, B. (2000) A short tandem repeat-based phylogeny for the human Y chromosome. *American Journal of Human Genetics*, **67**, 182.
- García-Obregón, S., Alfonso-Sánchez, M.A., Pérez-Miranda, A.M. & Peña, J.A. (2007) Polymorphic *Alu* insertions and the genetic structure of Iberian Basques. *Journal of Human Genetics*, **52**, 317.
- García-Obregón, S., Alfonso-Sánchez, M.A., Pérez-Miranda, A.M., Gómez-Pérez, L., de Pancorbo, M.M. & Peña, J.A. (2010) Genetic variability in autochthonous Basques from Guipuzcoa: a view from MHC microsatellites. *International Journal of Immunogenetics*, **37**, 279.
- Gómez-Pérez, L., Alfonso-Sánchez, M.A., Pérez-Miranda, A.M., García-Obregón, S., Builes, J.J., Bravo, M.L., De Pancorbo, M.M. & Peña, J.A. (2010) Genetic admixture estimates by *Alu* elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia. *Annals of Human Biology*, **37**, 488.
- Guo, S.W. & Thompson, E.A. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361.
- Hajeer, A.H. & Hutchinson, I.V. (2000) TNF- α gene polymorphism: clinical and biological implications. *Microscopy research and technique*, **50**, 216.
- Hearne, C.M., Ghosh, S. & Todd, J.A. (1992) Microsatellites for linkage analysis of genetic traits. *Trends in Genetics*, **8**, 288.
- Hedrick, P.W. (1982) Genetic hitchhiking: a new factor in evolution? *BioScience*, **32**, 845.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860.
- Kulski, J.K. & Dunn, D.S. (2005) Polymorphic *Alu* insertions within the major histocompatibility complex class I genomic region: a brief review. *Cytogenetic and Genome Research*, **110**, 193.
- Kulski, J.K., Martinez, P., Longman-Jacobsen, N., Wang, W., Williamson, J., Dawkins, R.L., Shiina, T., Naruse, T. & Inoko, H. (2001) The association between HLA-A alleles and an *Alu* dimorphism near HLA-G. *Journal of Molecular Evolution*, **53**, 114.
- Kulski, J.K., Dunn, D.S., Hui, J., Martinez, P., Romphruk, A.V., Leelayuwat, C., Tay, G.K., Oka, A. & Inoko, H. (2002) *Alu* polymorphism within the MICB gene and association with HLA-B alleles. *Immunogenetics*, **53**, 975.
- Lie, B.A. & Thorsby, E. (2005) Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Current Opinion in Immunology*, **17**, 526.
- Liu, K. & Muse, S. (2005) PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics*, **21**, 2128.
- Locksley, R.M., Killeen, N. & Lenardo, M.J. (2001) The TNF and TNF receptor review superfamilies: integrating mammalian biology. *Cell*, **104**, 487.
- Loisel, D.A., Alberts, S.C. & Ober, C. (2008) Functional significance of MHC variation in mate choice, reproductive outcome, and disease risk. in: *Evolution in Health and Disease* (eds Stearns, S.C. & Koella, J.C), p. 95. Publisher: Oxford Scholarship Online Monographs. Oxford University Press, New York.
- Mantullo, G., Griffo, R., Mangione, A., Cappello, N., Rendine, A. & Piazza, A. (1995) Analysis of HLA-A,C, B, DR, and DQ Loci in an Italian sample of possible celtic origin. *Tissue Antigens*, **45**, 295–301.
- McClelland, E.E., Penn, D.J. & Potts, W.K. (2003) Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, **71**, 2079.
- McManus, R., Moloney, M., Borton, M., Finch, A., Chuan, Y.T., Lawlor, E., Weir, D.G. & Kelleher, D. (1996) Association for celiac disease with microsatellite polymorphism close to the tumor necrosis factor genes. *Human Immunology*, **45**, 24.
- Meyer, D. & Thomson, G. (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*, **65**, 1.
- Pala, M., Achilli, A., Olivieri, A., Kashani, B.H., Perego, U.A., Sanna, D. *et al.* (2009) Mitochondrial haplogroup U5b3: a distant echo of the epipaleolithic in Italy and the legacy of the early Sardinians. *American Journal of Human Genetics*, **84**, 814.
- Peña, J.A., Calderón, R., Pérez-Miranda, A., Vidales, C., Dugoujon, J.M., Carrion, M. & Crouau-Roy, B. (2002) Microsatellite DNA markers from HLA region (D6S105, D6S265 and TNFa) in autochthonous basques from Northern Navarre (Spain). *Annals of Human Biology*, **29**, 176.
- Pérez-Miranda, A.M., Alfonso-Sánchez, M.A., Peña, J.A. & Calderón, R. (2003) HLA-DQA1 polymorphism in autochthonous Basques from Navarre (Spain): genetic position within European and Mediterranean scopes. *Tissue Antigens*, **61**, 465.
- Pérez-Miranda, A.M., Alfonso-Sánchez, M.A., Vidales, M.C., Calderón, R. & Peña, J.A. (2004) Genetic polymorphism and linkage disequilibrium of the HLA-DP region in Basques from Navarre (Spain). *Tissue Antigens*, **64**, 264.
- Piertney, S.B. & Oliver, M.K. (2006) The evolutionary ecology of the mayor histocompatibility complex. *Heredity*, **96**, 7.
- Ray, D.A., Walker, J.A., Hall, A., Llewellyn, B., Ballantyne, J., Christian, A.T., Turteltaub, K. & Batzer, M.A. (2005) Inference of human geographic origins using *Alu* insertion polymorphisms. *Forensic Science International*, **153**, 117.
- Reynolds, J., Weir, B.S. & Cockerman, C.C. (1983) Estimation of the coancestry coefficient: bases for a short term genetic distance. *Genetics*, **105**, 767.
- Rowold, D.J. & Herrera, R.J. (2005) On human STR sub-population structure. *Forensic Science International*, **151**, 59.
- Sanchez-Mazas, A. (2001) African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection. *Human Immunology*, **62**, 37.
- Segal, S. & Hill, A.V. (2003) Genetic susceptibility to infectious disease. *Trends in Microbiology*, **11**, 445.
- Shiina, T., Hosomichi, K., Inoko, H. & Kulski, J.K. (2009) The HLA genomic loci map: expression, interaction, diversity and disease. *Journal of Human Genetics*, **54**, 15.
- Takeuchi, F., Nabeta, H., Füssel, M., Conrad, K. & Frank, K.H. (2000) Association of the TNFa13 microsatellite with systemic sclerosis in Japanese patients. *Annals of Rheumatic Diseases*, **59**, 293.
- Tautz, D. & Schlötterer, C. (1994) Simple sequences. *Current Opinion in Genetics & Development*, **4**, 832.
- Terreros, M.C., Alfonso-Sánchez, M.A., Novick, G.E., Luis, J.R., Lacau, H., Lowery, R.K., Regueiro, M. & Herrera, R.J. (2009) Insights on human evolution: an analysis of *Alu* insertion polymorphisms. *Journal of Human Genetics*, **54**, 603.
- The MHC sequencing consortium. (1999) Complete sequence and gene map of a human mayor histocompatibility complex. *Nature*, **401**, 921.
- Thompson, S.D., Moroldo, M.B., Guyer, L., Ryan, M., Tombragel, E.M., Shear, E.S. *et al.* (2004) A genome-wide scan for juvenile rheumatoid arthritis in affected sibpair families provides evidence of linkage. *Arthritis Rheumatoid*, **50**, 2920.
- Tian, W., Wang, F., Cai, J.H. & Li, L.X. (2008) Polymorphic insertions in 5 *Alu* loci within the major histocompatibility

complex class I region and their linkage disequilibria with HLA alleles in four distinct populations in mainland China. *Tissue Antigens*, 72, 559.

- Törn, C., Hillman, M., Sanjeevi, C.B. & Landin-Olsson, M. (2006) Polymorphisms of TNF microsatellite marker and HLA-DR-DQ in diabetes mellitus—a study in 609 Swedish subjects. *Human Immunology*, 67, 527.
- Watterson, G.A. (1978) The homozygosity test of neutrality. *Genetics*, 88, 405.
- Yao, Y., Shi, L., Shi, L., Lin, K., Yu, L., Sun, H. *et al.* (2009) The association between HLA-A, -B alleles and major histocompatibility complex class I polymorphic *Alu* insertions in four populations in China. *Tissue Antigens*, 73, 575.
- Zhivotovsky, L.A., Underhill, P.A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T. *et al.* (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *American Journal of Human Genetics*, 74, 50.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1 Accession number (NCBI Entez Nucleotide), primers, annealing temperature (AT), and amplicon product sizes for the PCR amplification of 5 *Alu* loci from the human MHC region.

Table S2 UniSTS, repeat motif, HLA or HLA-like neighboring gene, primers sequences and PCR conditions for the analysis of five microsatellites from the human MHC region.

Table S3 Haplotypes based on five *Alu* markers (*Alu*-MicB, *Alu*TF, *Alu*HJ, *Alu*HG and *Alu*HF) and their frequencies in the populations used for phylogenetic analysis.

Table S4 Microsatellite allele frequencies in the populations used for phylogenetic analysis.

Table S5 Genotype frequencies for five MHC microsatellites in Valencia (Spain).

Table S6 Haplotypes based on 10 MHC markers (D6S291, D6S2666, D6S2792, *Alu*MicB, *Alu*TF, *Alu*HJ, D6S265, *Alu*HG, *Alu*HF and D6S105) and their frequencies in Valencia (Spain).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.