Coll. Antropol. **27** (2003) 2: 491–500
UDC 572.2:575.17(262)
Original scientific paper

Alu Insertions in the Iberian Peninsula and North West Africa – Genetic Boundaries or Melting Pot?

Emili González-Pérez[1], Marc Via[1], Esther Esteban[1], Antoni López-Alomar[1], Stéphane Mazieres[2], Nourdin Harich[3], Mostafa Kandil[3], Jean-Michel Dugoujon[2] and Pedro Moral[1]

[1] Department of Anthropology, Faculty of Biology, University of Barcelona, Barcelona, Spain
[2] Centre of Anthropology, University of Paul Sabatier, Toulouse, France
[3] Department of Biology, Faculty of Sciences, University of Chouaib-Doukkali, El Jadida, Morocco

A B S T R A C T

The Western Mediterranean Basin joins a set of ethnically different populations as Iberians and Basques in the North shore and Berbers and Arab-speakers in the South one. In spite of this differentiation, they have maintained historical contacts since ancient times. The existence of a possible common genetic background (specially for Berbers and Iberians) together with the genetic impact of the Islamic occupation of the Iberian Peninsula during 7 centuries are some of the intriguing anthropological questions that have been studied in this area using several classical and DNA markers. The aim of this work is to present the results on a survey of polymorphic Alu elements in 10 human populations of the Western Mediterranean. Recent Alu subfamilies include a significant number of polymorphic Alu insertions in humans. The polymorphic Alu elements are neutral genetic markers of identical descent with known ancestral states. This fact turns Alu insertions into useful markers for the study of human population genetics. A total number of 14 Alu insertions were analyzed in 5 Iberian populations, 3 Berber groups from North-Western Africa, an Arab-speaker population from Morocco and a sub-Saharan ethnic group from Ivory Coast. The results of this study allow the genetic characterization of Berber populations, which show a certain degree of differentiation from Arab-speaking groups of the same geographic area. Furthermore, a closer genetic distance between South Spain and Moroccan Berbers as compared with other Spanish samples supports a major genetic influx consistent with some (but not all) previous genetic studies on populations from the two shores of the Gibraltar Straits.

***Key words**: Alu elements, polymorphisms, Iberian populations, Berber populations*

Received for publication December 19, 2002

Introduction

Alu insertions are the most widely dispersed short interspersed nuclear elements (SINEs) representing more than 10% of the present human genome. Typically, an Alu insertion is a 300 bp long-sequence originated by dimeric evolution from the 7SL RNA gene, inserted into the genome through an intermediate RNA single strand generated by RNA polymerase III transcription.

Many of the Alu elements have been so recently retro transcribed that they are not fixed yet and appear to be polymorphics in the human genome[1–6]. According to diagnostic nucleotide changes on the original basic Alu sequence (master Alu element), the Alu insertions are classified into to 12 subfamilies that appeared in different times during the primate evolution and, hence, with different ages. Among the most recent of these subfamilies, Ya5/8 and Yb8 comprise a large number of polymorphic insertions that have been recently used in human population genetic studies. In contrast with other DNA markers often used in population studies, the extremely low probabilities of independent retro transposition and/or complete loss in the same exact genome site make the Alu insertions specially useful tools for detecting identity by descent and to long-term evolutionary reconstructions. Besides, the knowledge of the ancestral stage of the polymorphism allows the identification of the direction of evolutionary change and the possibility for rooting phylogenies. These features along with previous informative results turn Alu polymorphisms attractive and promising markers to the study of the genetic structure and historical reconstruction of human populations even at a micro geographical level[7,8].

Historical and demographic relationships between human populations from the Western Mediterranean basin are a highly interesting topic. Some studies tend to consider the Gibraltar Straits as an important genetic boundary to north-south population movements in the westernmost part of the Mediterranean, in agreement with a sharp and clear differentiation between North Africa and Iberian populations. These studies interpret this differentiation as related with the independent and parallel origin for northern and southern Mediterranean people from Neolithic migrations from the Middle East[9]. In contrast with this hypothesis, other opinions agree with a North African biological and cultural influence in the development of the Iberian Peninsula autochthonous populations as Iberians and Basques[14]. Any case, the potential role of Mesolithic North African Berbers in the historical development of Western Mediterranean populations is a matter of discussion. This possible influence should be traced back more than 5 thousand years when the aridification of the Sahara desert favored the movement of the ancient north African populations towards the Mediterranean coast and possibly contacting with north Mediterranean people. This ancient influence together with the also discussed impact of the historical Muslim domination of the Iberian Peninsula during eight centuries (8^{th}–15^{th}), might be genetically detected on the basis it was really considerable.

Recent studies have focused the genetic differentiation across the Gibraltar Straits leading to controversial results. A recent genetic study on the distribution of 11 Alu polymorphisms was interpreted as consistent with a north-south genetic differentiation in Western Mediterranean stressing the importance of the Gibraltar Straits as genetic boundary[11].

This paper deals too with data on Alu population variation in the same geographical area, using a larger number of different markers, yielding discrepant results. The distribution of 14 Alu inser-

tions has been scored in 10 populations from the Iberian Peninsula and North Africa, including also a sub-Saharan ethnic group from Ivory Coast as external reference. This information is applied to assess the population relationships between Western Mediterranean groups and clarify the divergences between North-West African populations, completing the general picture of the variation of this kind of polymorphisms in this geographical area.

Materials and Methods

A total of 1,126 individuals coming from different population groups were analyzed. Each population sample includes unrelated healthy blood donors whose four grandparents are natives from the same region. Informed consent was obtained from all subjects included in the study. The Western Mediterranean samples included in the study are listed in Figure 1. Besides, a total number of 122 individuals from a sub-Saharan sample (Ivory Coast) were included in the analysis that was used as external reference.

Genomic DNA was extracted from blood by classical phenol-chloroform method. Fourteen human-specific Alu polymorphic elements (CD4, TPA25, APO, ACE, Yb8NBC120, Yb8NBC125, B65, D1, FXIIIB, A25, PV92, HS2.43, Ya8NBC3 and Sb19.12) were genotyped in each sample by using the primers described previously[1,4,16,18,19]. The PCR amplification conditions for the first six loci were performed as described previously[5,19]. Four multiplex PCR procedures were used to test eight loci (HS2.43, Ya8NBC3, Sb19.12, B65, D1, FXIIIB, A25 and PV92) using primers described before[1,4,16,18,19].

Allele frequencies were computed by direct counting and Hardy-Weinberg equilibrium was checked by an exact test[20]. Gene diversity by population and locus was calculated according to the Nei's formula[21]. Geographical structure of the allele frequency variance was tested by a hierarchical analysis of the molecular variation[22] using F-Wright statistics from population clustering according to geographical criteria. Estimates were obtained using the Arlequin 2.000 computer package[23].

Fst-related genetic distances were calculated between pairs of populations[24] and represented in a neighbor joining tree[25] using the PHYLIP 3.6 package[26]. The topology of the tree was assessed through 1,000 bootstrap iterations. The genetic relationships between the examined populations were also depicted by principal components analysis (PCA).

The Delaunay network analysis was used to identify principal boundaries or regions of sharp genetic change[9]. With this intention, we defined a subset of pairs of contiguous populations connected by 13 edges. The genetic distances between each pair of samples were allocated in each edge, and the high genetic distances were joined to trace the principal genetic boundary in the region.

The isolation degree between populations and hence, the sense of defining genetic boundaries, was tested by the ISOLDE program in the Genepop package[27], that elucidate if the observed differences could be attributable to an isolation by distance or to sharp geographical discontinuities preventing human migrations. Finally, gene flow between Moroccan Berber populations and South Iberian was calculated using the ADMIX 1.0 program[28].

Results and Discussion

The pattern of Alu insertion frequency distribution for the 14 loci examined in the 10 populations typed is shown in Figure 2. Fifteen out of 140 tests of Hardy-Weinberg equilibrium show a significant departure from equilibrium ($p < 0.01$).

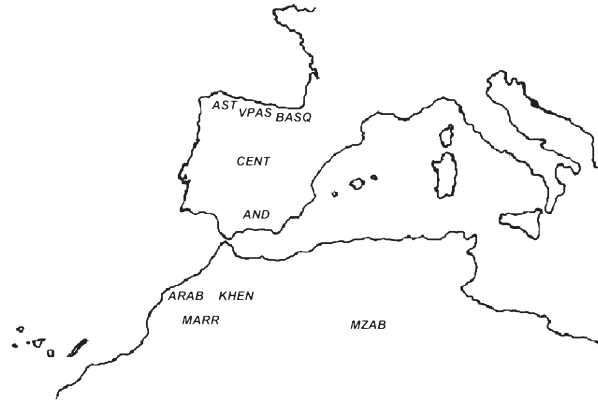


Fig. 1. Map of the Western Mediterranean Basin with the locations of the studied populations: ASTU – Asturias; VPAS – Pas Valley; BASQ – Basques; CENT – Center of Spain; AND – Andalusians; KHEN – Moroccan Berbers from Khénifra; MARR – Moroccan Berbers from Marrakech; ARAB – Arab-speakers from Morocco (El-Jadida); MZAB – Algerian Berbers (Mozabites).

After application of Bonferroni correction, only six comparisons (4.2%) maintain significant deviations. These departures probably reflect random statistical fluctuations. Heterozygosities across loci and population are shown in Table 1. The 14 Alu polymorphisms show significant differences in their heterozygosities (Kruskal-Wallis test, p=0.000) depending on the allele frequencies of each polymorphism. Mean heterozygosities by locus ranged from 0.1745 (APO A1) to near 0.5 (B65 and TPA25). Focusing on the heterozygosity by population, no significant

TABLE 1
AVERAGE GENE DIVERSITY (HETEROZYGOSITIES) BY LOCUS AND POPULATION

Locus		Population	
A25	0.2577±0.031	Basques	0.3674±0.039
ACE	0.4646±0.016	Asturias	0.3825±0.041
APOA1	0.1745±0.036	Pas Valley	0.3621±0.041
B65	0.4864±0.007	Andalusians	0.3621±0.037
CD4	0.4051±0.023	Center Spain	0.3664±0.037
D1	0.4391±0.018	Arab Morocco	0.2953±0.046
FXIIIB	0.4622±0.014	Ber-Khenifra	0.3814±0.036
HS2.43	0.1031±0.021	Ber-Marrakech	0.3365±0.035
PV92	0.2607±0.046	Ber-Mzab	0,3433±0.043
Sb19.12	0.3781±0.033	Ivory Coast	0.3580±0.041
TPA25	0.4807±0.014		
Ya8NBC3	0.3882±0.023		
Yb8NBC120	0.4540±0.017		
Yb8NBC125	0.2176±0.034		

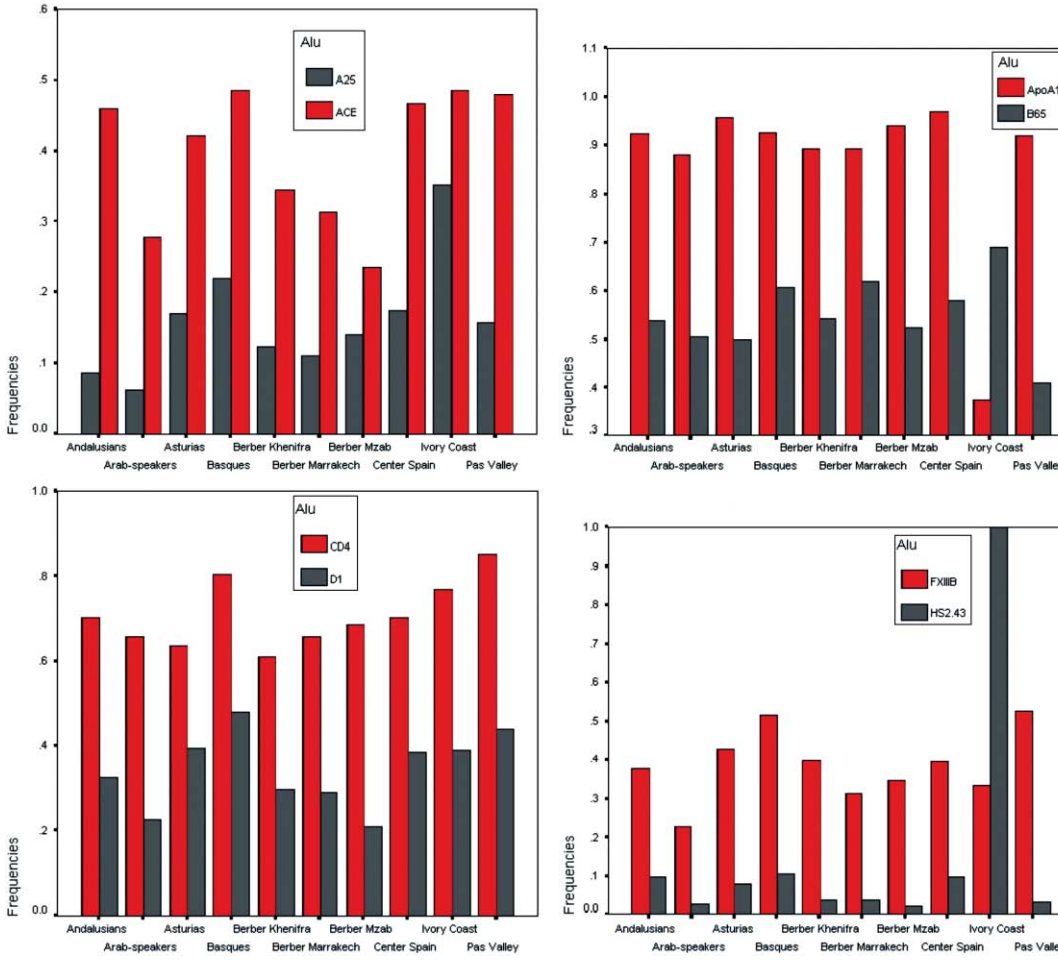
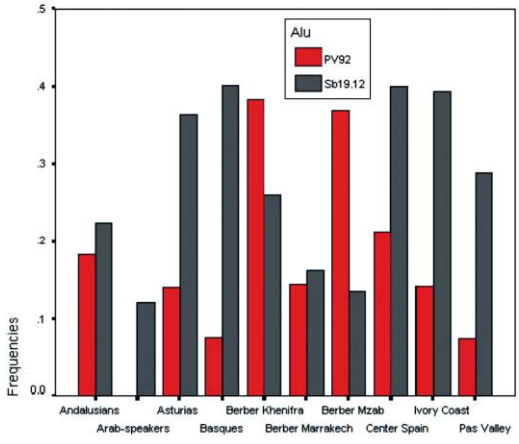
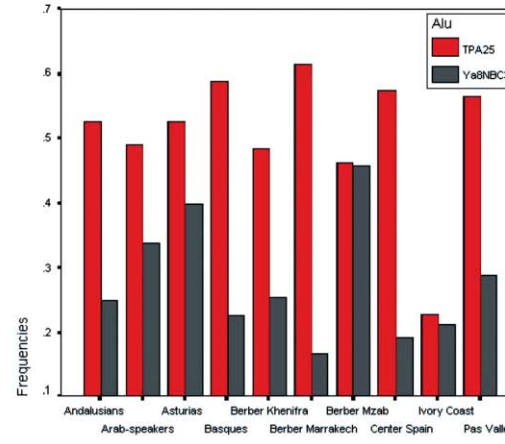


Fig. 2. Allele (+) frequencies for the 14 Alu polymorphic markers analyzed.
(continued on next page)

differences between samples are found (Kruskal-Wallis test, p=0.844). The average population heterozygosity across loci is relatively high in all cases (from 0.2953 in Arabs to 0.3825 in Asturians) taking into account the biallelic features of the examined markers.

As for population relationships, Reynolds genetic distances between Western Mediterranean samples were obtained and represented in a Neighbor-Joining tree (Figure 3). The tree does not show the clear separation between Iberian and North-African populations expected from other published results[11]. In contrast, the populations are distributed in a central cluster that includes the majority of Iberian and Moroccan groups. As more differentiated, the extreme positions correspond to the Pas Valley Spanish sample while the Moroccan Arab-speaking and Mozabites Algerian groups are placed on





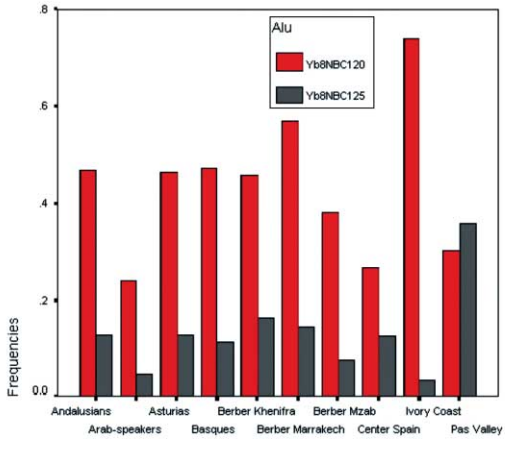


Fig. 2. (continued from page before)

the opposite side of the tree. Low bootstrap values (lower than 50% after 1,000 replicates) correspond to branches connecting Moroccan Berbers (Khénifra and Marrakech) and South Iberia (Andalusia) showing the lack of a clear differentiation among these populations.

The results of a principal component analysis (PCA) are shown in Figure 4. The first two principal components account for 66.32% of the variance observed (36.01 and 30.31% respectively). Although the sub-Saharan group was included in this analysis, the general picture of Mediterranean populations is very similar to that from the NJ tree. The population distribution along the first PC underlines the extreme positions of North Spaniards (Basques and Pas Valley) and Arab-speakers, as well as the genetic similarity between Moroccan Berbers and South Iberia in the central position. This distribution may be mainly attributed to the frequencies of Sb19.12, ACE and D1 Alu polymorphisms (the correlations with this axis were > 87%). The second PC differentiates the Sub-Saharan sample from the

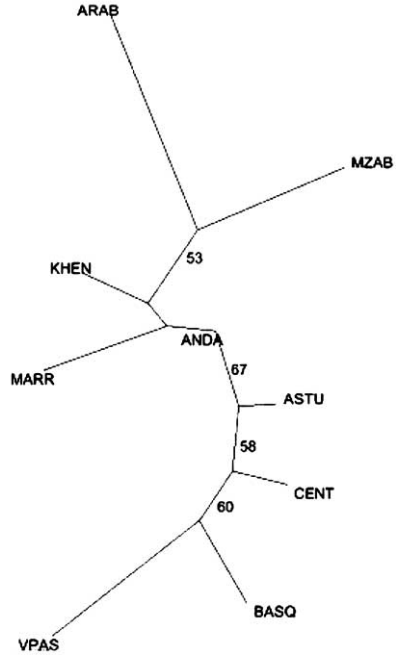


Fig. 3. Neighbor-joining tree showing relationships among the nine western Mediterranean samples for the 14 Alu polymorphic elements analysed. ASTU – Asturias; VPAS – Pas Valley; BASQ – Basques; CENT – Center of Spain; ANDA – Andalusians; KHEN – Moroccan Berbers from Khénifra; MARR – Moroccan Berbers from Marrakech; ARAB – Arab-speakers from Morocco (El-Jadida); MZAB – Algerian Berbers (Mozabites).

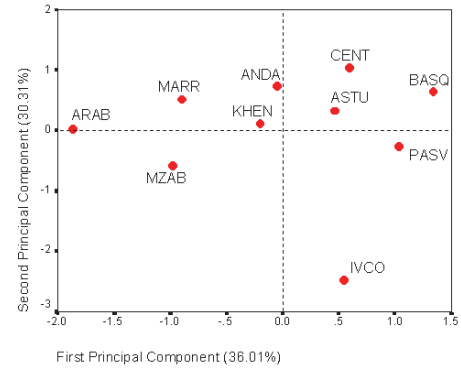


Fig. 4. Principal Components Analysis (PCA) of the allele frequencies at 14 Alu polymorphic loci in North Africa and the Iberian Peninsula. ASTU – Asturias; VPAS – Pas Valley; BASQ – Basques; CENT – Center of Spain; ANDA – Andalusians; KHEN – Moroccan Berbers from Khénifra; MARR – Moroccan Berbers from Marrakech; ARAB – Arab-speakers from Morocco (El-Jadida); MZAB – Algerian Berbers (Mozabites); IVCO – Ivory Coast (Ahizi).

Mediterranean's, and this differentiation is mainly associated with HS2.43, APO and TPA25 Alu markers.

A hierarchical AMOVA analysis between the geographical groups (Iberian and the NW African regions), indicates that a greater part of the total allele frequency variance (around 4%) may be attributed to the variation within geographical regions (2.3%) rather than to the between-region variability (1.8%), failing to evidence any particular and/or relevant genetic differentiation between populations settled in both sides of the Gibraltar Straits.

The genetic boundary analysis through the Delaunay method (Figure 5) shows that the most important barrier separates Southern Iberian together with Moroccan Berbers (Khénifra and Marrakech) from the remaining North African populations (Arab-speakers from Morocco and Algerian Berbers). This result is consistent with the presence of significant gene flow between South Iberia and Morocco.

As an indirect way to test the consistency of clear genetic boundaries from the observed Alu variation, the correlation between geographical and genetic distances was tested under the isolation by distance model using the ISOLDE program[27]. A highly significant correlation (p=0.001, calculated after 1,000 bootstrap iterations) was found suggesting an important role of the geographic distance for the interpretation of the genetic variability in Western Mediterranean. These results are hardly consistent with the existence of

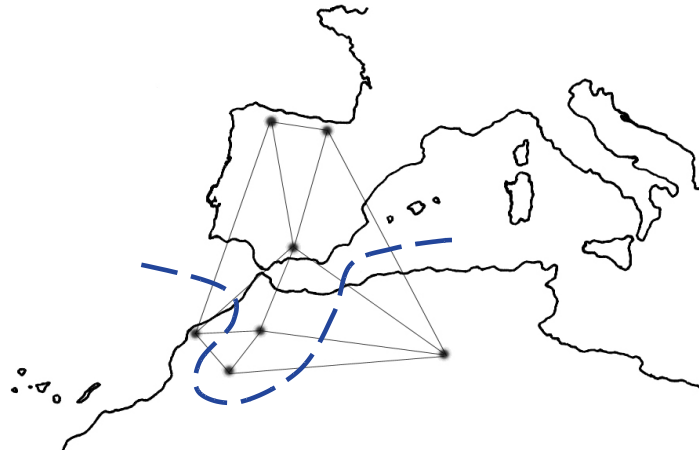


Fig. 5. First genetic boundary (dashed line) recognized on the basis of genetic distances after a Delaunay network analysis among geographical locations of the samples analyzed.

strong genetic barriers as diffe- rentiation mechanisms in this geographical region.

From our data, the North African gene flow into Southern Iberian was estimated around 50.9% (±11.6%) using the Bertorelle and Excoffier[29] method and Asturians (in North Spain) and Khénifra Moroccan Berbers as parental populations. When Arab-speaking Moroccans instead Berbers were used as parental population, the amount of genetic flow was around 28% (±9.1%). Although the amount of gene flow is variable as expected on the basis of populations taken as parental, our data indicate a substantial gene flow between populations living in both shores of the Gibraltar Straits.

Results in this work contrast with previous studies based on the analysis of the same kind of genetic markers (Alu polymorphisms). This apparent contradiction could be explained if the heterogeneity between populations within each geographical region (NW Africa and Iberian Peninsula) is greater than differences between the regions. In this case, the larger number of markers and, likely also, of sample sizes, the lesser possible population dependence due to the intraregional variability.

In conclusion, the data presented in this study on the variation of 14 Alu polymorphisms provide new data for a more complete definition of the Western Mediterranean populations, allowing the characterization of the Moroccan Berber populations among the Western Mediterranean groups. The frequency variation of the markers analyzed fit well with an isolation by distance model instead of sharp geographical discontinuities preventing human migrations over the Gibraltar Straits. In short, our data support significant gene flow through the Gibraltar Straits, consistent with historical and biological contacts between these Berber groups and the Southern Spain, in contrast with previous results (Comas et al 2000). These discrepancies at least could be indicating that not all Iberian populations are so different from their North African neighbors. At least a part of this biological influence might be related to the Muslim conquest of the Iberian Peninsula from 8th to 15th centuries.

Acknowledgements

This work was supported by the Dirección General de Investigación Científica y Técnica from Spain (PB98-1235-C3-01), the Comissionat per a Universitats i Recerca de la Generalitat de Catalunya (2000 SGR00033) and the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (2001FI 00177 UB and 2002FI 00516 grant).

REFERENCES

1. BATZER, M. A., P. L. DEININGER, Genomics, 9 (1991) 481. — 2. BATZER, M. A., M. STONEKING, M. ALEGRIA-HARTMAN, H. BAZAN, D. H. KASS, T. H. SHAIKH, G. E. NOVICK, P. A. IOANNOU, W. D. SCHEER, R. J. HERRERA, P. L. DEININGER, Proc. Natl. Acad. Sci. USA, 91 (1994) 12288. — 3. BATZER, M. A., S. S. ARCOT, J. W. PHINNEY, M. ALEGRIA-HARTMAN, D. H. KASS, S. M. MILLIGAN, C. KIMPTON, P. GILL, M. HOCHMEISTER, P. A. IOANNOU, R. J. HERRERA, D. A. BOUDREAU, W. D. SCHEER, B. J. KEATS, P. L. DEININGER, M. STONEKING, J. Mol. Evol., 42 (1996) 22. — 4. ARCOT, S. S., A. W. ADAMSON, J. E. LAMERDIN, B. KANAGY, P. L. DEININGER, A. V. CARRANO, M. A. BATZER, Genome Res., 6 (1996) 1084. — 5. STONEKING, M., J. J. FONTIUS, S. L. CLIFFORD, H. SOODYALL, S. S. ARCOT, N. SAHA, T. JENKINS, M. A. TAHIR, P. L. DEININGER, M. A. BATZER, Genome Res., 7 (1997) 1061. — 6. NOVICK, G. E., C. C. NOVICK, J. YUNIS, E. YUNIS, P. ANTÚNEZ DE MAYOLO, W. D. SCHEER, P. L. DEININGER, M. STONEKING, D. S. YORK, M. A. BATZER, R. J. HERRERA, Hum. Biol., 70 (1998) 23. — 7. MORAL, P., M. BAO, E. GONZÁLEZ-PÉREZ, A. LÓPEZ-ALOMAR, L. VARESI, M. MEMMI, G. VONA, Antropologia Contemporanea, (2000) 77. — 8. GONZÁLEZ-PÉREZ, E., A. LÓPEZ-ALOMAR, M. BAO, M. VIA, E. ESTEBAN, N. VALVENY, N. HARICH, M. KANDIL, J. M. DUGOUJON, P. MORAL (XIIth SEAB Congress, Barcelona, 2001). — 9. SIMONI, L., P. GUERESI, D. PETTENER, G. BARBUJANI, Hum. Biol., 71 (1999) 399. — 10. ARNAIZ-VILLENA, A., J. MARTÍNEZ-LASO, E. GÓMEZ-CASADO, N. DÍAZ-CAMPOS, P. SANTOS, A. MARTINHO, H. BREDA-COIMPBRA, Immunogenetics, 47 (1997) 37. — 11. COMAS, D., F. CALAFELL, N. BENCHEMSI, A. HELAL, G. LEFRANC, M. STONEKING, M. A. BATZER, J. BERTRANPETIT, A. SAJANTILA, Hum. Genet., 107 (2000) 312. — 12. BOSCH, E., F. CALAFELL, A. PÉREZ-LEZAUN, J. CLARIMÓN, D. COMAS, E. MATEU, R. MARTÍNEZ-ARIAS, B. MORENA, Z BRAZEK, O. AKHAYAT, A. SEFIANI, G. HARITI, A. CAMBON-THOMSEN, J. BERTRANPETIT, Eur. J. Hum. Genet., 8 (2000) 360. — 13. RANDO, J. C., F. PINTO, A. M. GONZÁLEZ, M. HERNÁNDEZ, J. M. LARRUGA, V. M. CABRERA, H. J. BANDELT, Ann. Hum. Genet., 62 (1998) 531. — 14. GÓMEZ-CASADO, E., P. MORAL, J. MARTÍNEZ-LASO, A. GARCÍA-GÓMEZ, L. ALLENDE, C. SILVERA-REDONDO, J. LONGAS, M. GONZÁLEZ-HEVILLA, M. KANDIL, J. ZAMORA, A. ARNAIZ-VILLENA, Tissue Antigens, 55 (2000) 239. — 15. FLORES, C., N. MACA-MEYER, A. M. GONZÁLEZ, V. M. CABRERA, Ann. Hum. Genet., 64 (2000) 321. — 16. ARCOT, S. S., J. J. FONTIUS, P. L. DEININGER, M. A. BATZER, Biochim. Biophys. Acta, 1263 (1995) 99. — 17. ROY-ENGEL, A. M., M. L. CARROLL, E. VOGEL, R. K. GARBER, S. V. NGUYEN, A. H. SALEM, M. A. BATZER, P. L. DEININGER, Genetics, 159 (2001) 279. — 18. WATKINS, W. S., C. E. RICKER, M. J. BAMSHAD, M. L. CARROLL, S. V. NGUYEN, M. A. BATZER, H. C. HARPENDING, A. R. ROGERS, L. B. JORDE, Am. J. Hum. Genet., 68 (2001) 738. — 19. EDWARDS, M. C., R. A. GIBBS, Genomics, 14 (1992) 590. — 20. GUO, S., E. THOMPSON, Biometrics, 48 (1992) 361. — 21. SAITOU, N., M. NEI, Mol. Biol. Evol., 4 (1987) 406. — 22. EXCOFFIER, L., P. E. SMOUSE, J. M. QUATRO, Genetics, 131 (1992) 479. — 23. SCHNEIDER, S., D. ROESSLI, L. EXCOFFIER: Arlequin ver. 2000: A software for population genetics data analysis. (Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000). — 24. REYNOLDS, J., B. S. WEIR, C. C. COCKERMAN, Genetics, 105 (1983) 767. — 25. SAITOU, N., M. NEI, Mol. Biol. Evol., 4 (1987) 406. — 26. FELSENSTEIN, J., Cladistics, 5 (1989) 164. — 27. RAYMOND, M., F. ROUSSET, J. Heredity, 86 (1995) 248. — 28. ROBERTS, D. F., R. W. HIORNS, Hum. Biol., 37 (1965) 38. — 29. BERTORELLE, G., L. EXCOFFIER, Mol. Biol. Evol., 15 (1998) 1298.

E. González-Pérez

Department of Anthropology, Faculty of Biology, University of Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain

ALU INSERCIJE NA IBERIJSKOM POLUOTOKU I U SJEVEROZAPADNOJ AFRICI – GENETIČKE BARIJERE ILI LONAC ZA TALJENJE?

S A Ž E T A K

Zapadni mediteranski bazen povezuje skupinu etnički različitih populacija kao što su Iberijci i Baski na sjevernoj te Berberi i Arapi na južnoj obali. Unatoč ovoj diferencijaciji, tijekom povijesti ovi narodi još od pradavnih vremena održali su kontakte. Postojanje mogućih zajedničke genetske podloge (osobito Berbera i Iberijaca) zajedno s genetskim utjecajem islamske okupacije iberijskog poluotoka tijekom 7 stoljeća neka su od intrigantnih antropoloških pitanja koja su istraživanja u ovom području korištenjem nekoliko klasičnih i DNK markera. Cilj ovog rada bio je prikazati rezultate istraživanja polimorfnih Alu elemenata u 10 ljudskih populacija zapadnog Mediterana. U ljudi, skorašnje Alu pod-obitelji uključuju značajan broj polimorfnih Alu insercija. Polimorfni Alu elementi su neutralni genetski markeri identičnog porijekla s poznatim ancestralnim stanjem. Ova činjenica pretvara Alu insercije u korisne markere za istraživanja genetike ljudskih populacija. Analizirano je ukupno 14 Alu insercija u 5 iberijskih populacija, 3 berberske skupine iz sjeverozapadne Afrike, jedna populacije arapskih govornika iz Maroka, te jedna subsaharska etnička skupina iz obale bjelokosti. Rezultati ove studije dopuštaju genetičku karakterizaciju berberske populacije, koja pokazuje određeni stupanj diferencijacije od skupine arapskih govornika istog zemljopisnog područja. Štoviše, bliža genetska distanca između južne Španjolske i Marokanskih Berbera u usporedbi s drugim uzorcima iz Španjolske govori u prilog snažnog genetskog utjecaja koji je konzistentan (no ne u cijelosti) s nekim prethodnim genetskim studijama populacija dvaju obala Gibraltarskih vrata.