Variació genètica i evolució d'elements *Alu* recents en poblacions humanes. Inferències biodemogràfiques i filogeogràfiques

Genetic variation and evolution of recent *Alu* elements in human populations. Biodemographic and philogeographic inferences

Memòria presentada per

Emili González Pérez

per optar al grau de
Doctor per la Universitat de Barcelona

Dirigida pel Dr. Pedro Moral Castrillo, Professor Titular d'Antropologia Física
de la Unitat d'Antropologia del Departament de Biologia Animal
de la Universitat de Barcelona.

Programa de Doctorat d'*Antropologia Biològica*, bienni 1999-2001.
Departament de Biologia Animal – Facultat de Biologia

Pedro Moral Castrillo	Clara García Moro	Emili González Pérez
Director	*Tutora*	*Doctorand*

Resultats

Water, water, every where,
Nor any drop to drink.

Samuel Taylor Coleridge, *Rime of the Ancient Mariner*

7 Informe del Director de Tesi sobre els Articles

La tesi doctoral "Variació genètica i evolució d'elements *Alu* recents en poblacions humanes. Inferències biodemogràfiques i filogeogràfiques" està articulada sobre els resultats originals obtinguts per Emili González Pérez i que han estat publicats en cinc revistes internacionals diferents, totes elles sota el sistema de revisió *peer review*.

El fil conductor de la recerca realitzada i de les cinc publicacions presentades en la tesi és l'anàlisi de la variació poblacional de marcadors genètics de tipus inserció *Alu* i d'una sèrie de STR's lligats a les mateixes, així com la seva aplicació a la reconstrucció de la història demogràfica de diferents grups humans actuals. La gran quantitat de dades genètiques poblacionals obtingudes (tant pel que fa als marcadors com a les poblacions estudiades) i el seu tractament han suposat un importat avenç en el coneixement de la variació genètica actual en humans i en la comprensió dels seus orígens.

La importància de la recerca realitzada en els camps de l'Antropologia Biològica i de la Genètica de les Poblacions Humanes queda demostrada per la qualitat de les revistes on han estat publicats els resultats obtinguts. Sobre aquest respecte, cal mencionar que les revistes:

- *Collegium Antropologicum* està indexada en el SCI (*Science Citation Index*) amb un factor d'impacte actual (i creixent en els darrers anys) de 0,414 i es troba classificada en el segon quartil de l'àrea "Anthropology" (posició 31/53);

- *Annals of Human Biology* és la revista de la *Society for the Study of Human Biology*, està indexada en el SCI amb un factor d'impacte actual de 1,060 i es troba situada en el segon quartil de l'àrea "Biology" (posició 37/64);

- *Journal of Human Genetics* és una revista associada a la *Japan Society of Human Genetics*, està indexada en el SCI amb un factor d'impacte actual de 2,275 i es troba en el segon quartil de l'àrea "Genetics & Heredity" (posició 80/132);

- *American Journal of Physical Anthropology* és la revista oficial de la *American Association of Physical Anthropologists*, està indexada en el SCI y SSCI amb un factor d'impacte de 2,273 i es troba classificada en les primeres posicions del primer quartil de l'àrea "Anthropology" (posició 3/53);

- *Annals of Human Genetics* és una revista clàssica sobre la biologia de la variació humana i editada pel *University College London*, està indexada en el SCI amb un factor d'impacte en el moment de la publicació de 2,727 i situada en el tercer quartil de l'àrea "Genetics & Heredity" (posició 59/131), tot i que actualment es situa en el segon quartil de l'àrea (posició 84/138).

Signat: Dr. Pedro Moral Castrillo
Barcelona, 29 de setembre de 2009

8 Resultats I:

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

Emili González-Pérez, Marc Via, Esther Esteban, Antoni López-Alomar, Stéphane Mazières, Nourdin Harich, Mostafa Kandil, Jean-Michel Dugoujon & Pedro Moral.

Collegium Antropologicum 2003; 27(2): 491-500

8.1 Insercions Alu *a la península Ibèrica i al Nord-Oest d'Àfrica: barreres genètiques o gresol de poblacions?*

Aquest és el primer dels manuscrits que directament van ser obtinguts del treball de recerca programat en el projecte inicial per a la consecució d'aquesta tesi doctoral. L'article està plantejat com una primera aproximació parcial, a partir de la variació genètica associada als elements *Alu*, a les relacions poblacionals a la regió mediterrània, i més particularment, a les referides a l'extrem més occidental d'aquest territori i al paper de poblacions particulars com els berbers en la configuració i evolució del panorama genètic actual a la regió.

La conca mediterrània occidental presenta un conjunt de característiques molt interessants per a l'abordatge i l'estudi de la genètica de poblacions humanes. Així, per exemple, algunes de les seves poblacions s'han considerat com algunes de les ètnicament més diferenciades dins la diversitat eurasiàtica actual. Entre aquests casos, cal destacar grups humans com els bascos i els pasiegos al nord-oest mediterrani i els berbers i arabòfons a la riba sud del mateix mar. Tot i aquestes particulars diferenciacions, es conserven evidències de contactes poblacionals més o menys intensos en el decurs dels segles d'història de poblament humà del territori.

Alguns dels aspectes més estudiats en aquest àmbit, a partir de marcadors clàssics i genètics com el mtDNA i el cromosoma Y, han estat el grau de *background* comú que podrien compartir poblacions ancestrals dels ibers i els berberòfons actuals així com el possible impacte que hagués tingut la ocupació musulmana de la península Ibèrica durant més de set segles.

L'objectiu primordial d'aquest treball va ser presentar una base de dades suficientment àmplia i completa de marcadors d'inserció *Alu* polimòrfics en 10 poblacions antropològicament ben caracteritzades de la regió més occidental del Mediterrani, a ambdues ribes de l'estret de Gibraltar.

Les característiques dels elements *Alu* polimòrfics, marcadors genètics neutres amb estat ancestral conegut i que permeten inferir identitat per descendència, van fer considerar l'anàlisi en detall de la seva variabilitat genètica com una eina especialment útil en la reconstrucció de les relacions poblacionals existents entre aquestes poblacions particulars i per a intentar inferir certs aspectes de la seva història particular.

D'aquesta manera, els 14 elements *Alu* inicialment estudiats en el projecte van ser genotipats en cinc poblacions representatives de la diversitat humana present a la península Ibèrica, en tres poblacions berberòfones del nord-oest africà, en una població característicament arabòfona del Marroc i en un grup poblacional extern a la regió (originari de Costa d'Ivori).

Els resultats del treball van permetre una primera caracterització dels grups berbers del Marroc, que mostrava un cert grau de diferenciació respecte a la població arabòfona de la mateixa àrea geogràfica. A més a més, l'estudi va poder observar una relativa proximitat biològica dels grups berbers marroquins amb la població del sud peninsular, més marcada que no pas amb la resta de poblacions ibèriques i amb el doble d'influència dels grups berberòfons que d'arabòfons al sud d'Espanya.

Aquestes observacions, en conjunt, donen suport a una diferenciació principal en base a la distància geogràfica, així com a un considerable grau de flux gènic a través de l'estret de Gibraltar durant la història d'aquests grups humans, un fet que també ha estat observat en alguns –però no en tots– estudis previs basats en la caracterització genètica dels grups humans a ambdues ribes de l'estret. Aquesta darrera dada, estaria associada, almenys en part, a la conquesta musulmana de la península Ibèrica i la seva permanència al sud de la mateixa entre els segles VIII i XV.

8.2 Informe del director sobre la participació del doctorand

El Dr. **Pedro Moral Castrillo**, Professor Titular del Departament de Biologia Animal de la Universitat de Barcelona i director de la tesi doctoral **"Variació genètica i evolució d'elements *Alu* recents en poblacions humanes. Inferències biodemo-gràfiques i filogeogràfiques"** presentada pel doctorand **Emili González Pérez**, desitja fer constar que la participació del doctorand en l'elaboració de l'article **"*Alu* Insertions in the Iberian Peninsula and North West Africa: Genetic Boundaries or Melting Pot?"** publicada per la revista *Collegium Antropologicum*, ha consistit en les següents tasques principals:

- Processament de mostres: extracció de DNA de les diverses mostres poblacionals utilitzades en el treball (100 %)
- Disseny del treball i selecció de marcadors analitzats, conjuntament amb el Dr. Pedro Moral
- Genotipatge al laboratori de marcadors d'inserció *Alu* (100 %)
- Elaboració de les bases de dades de resultats (100 %)
- Anàlisi estadística dels resultats
- Redacció de l'article, conjuntament amb el Dr. Pedro Moral

Complementàriament, cal assenyalar que cap dels coautors d'aquest article ha utilitzat, ni implícitament ni explícitament, els resultats d'aquest treball per a l'elaboració d'una altra tesi doctoral. En conseqüència, aquest article forma part, de manera exclusiva, del treball de recerca en el que s'emmarca la tesi del doctorand Emili González Pérez.

Signat: Dr. Pedro Moral Castrillo
Barcelona, 29 de setembre de 2009

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 (8th –15th), 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 ($p < 0.01$).

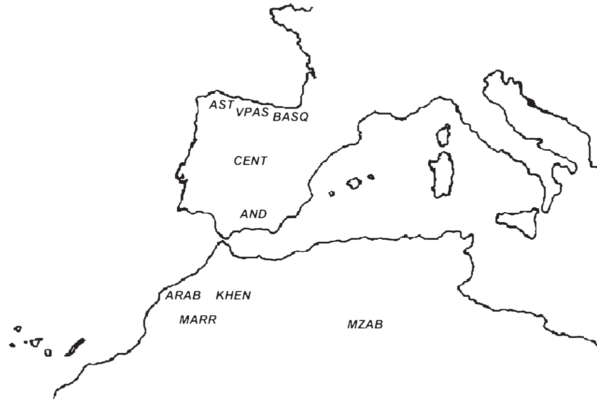

Fig. 1. Map of the Western Mediterranean Basin with the locations of the studied populations: ASTU – Asturias; VPAS – Pas Valley; BASQ – Basques; CENT – Center of Spain; AND – Andalusians; KHEN – Moroccan Berbers from Khénifra; MARR – Moroccan Berbers from Marrakech; ARAB – Arab-speakers from Morocco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

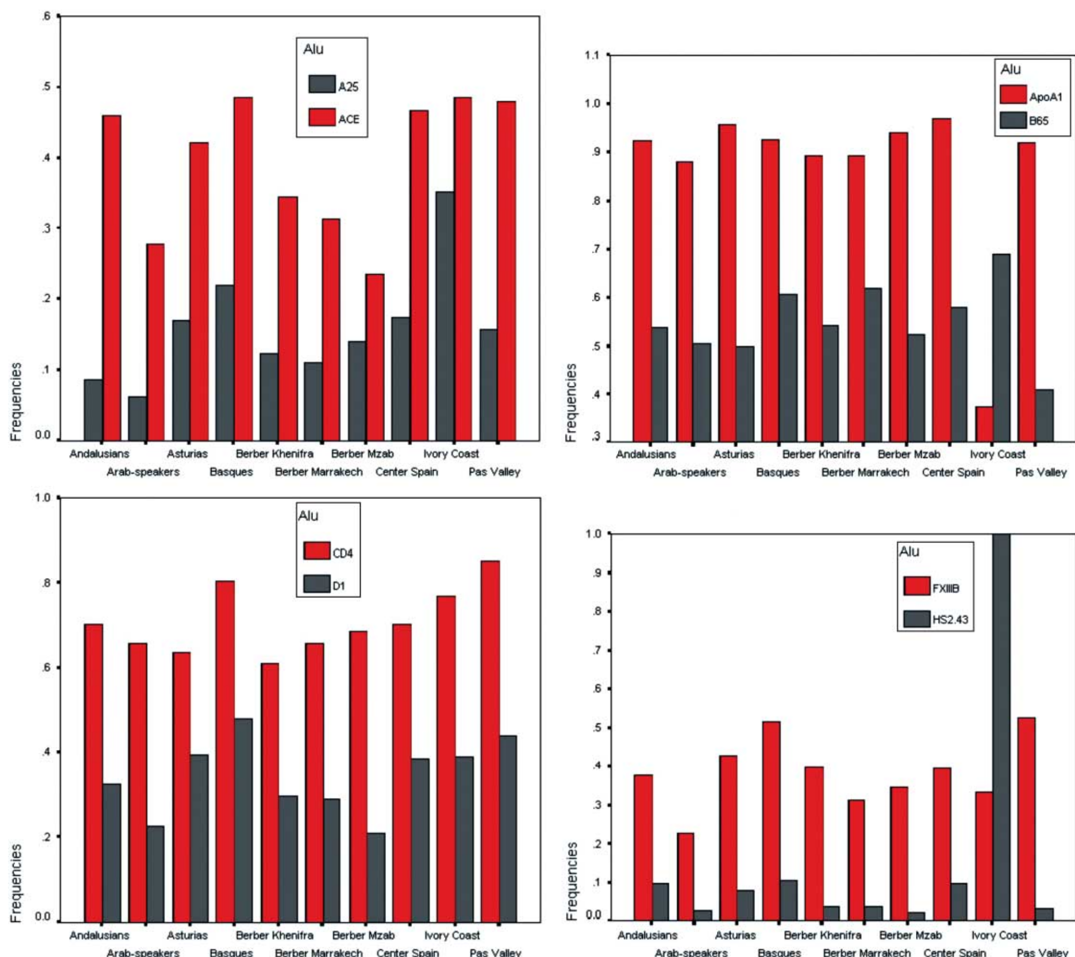

Fig. 2. Allele (+) frequencies for the 14 Alu polymorphic markers analyzed.
(continued on next page)

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

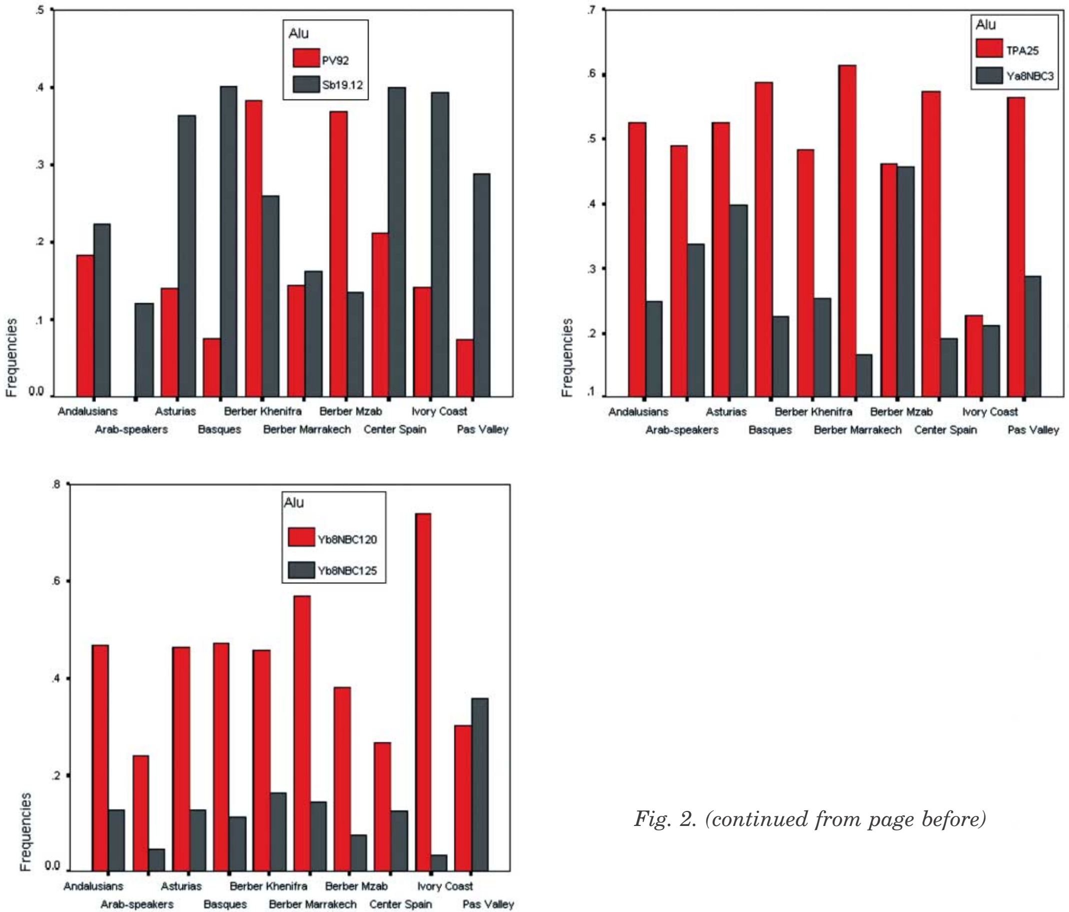

Fig. 2. (continued from page before)

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

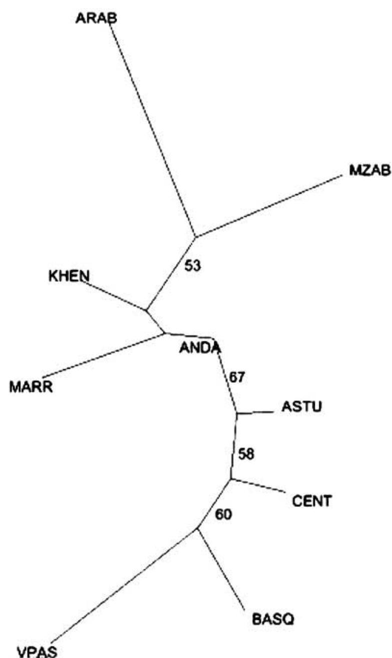

Fig. 3. Neighbor-joining tree showing relationships among the nine western Mediterranean samples for the 14 Alu polymorphic elements analysed. ASTU – Asturias; VPAS – Pas Valley; BASQ – Basques; CENT – Center of Spain; ANDA – Andalusians; KHEN – Moroccan Berbers from Khénifra; MARR – Moroccan Berbers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

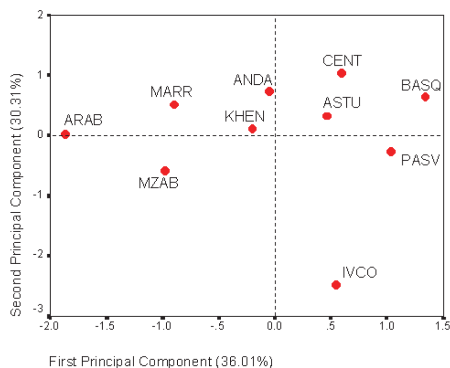

Fig. 4. Principal Components Analysis (PCA) of the allele frequencies at 14 Alu polymorphic loci in North Africa and the Iberian Peninsula. ASTU – Asturias; VPAS – Pas Valley; BASQ – Basques; CENT – Center of Spain; ANDA – Andalusians; KHEN – Moroccan Berbers from Khénifra; MARR – Moroccan Berbers from 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

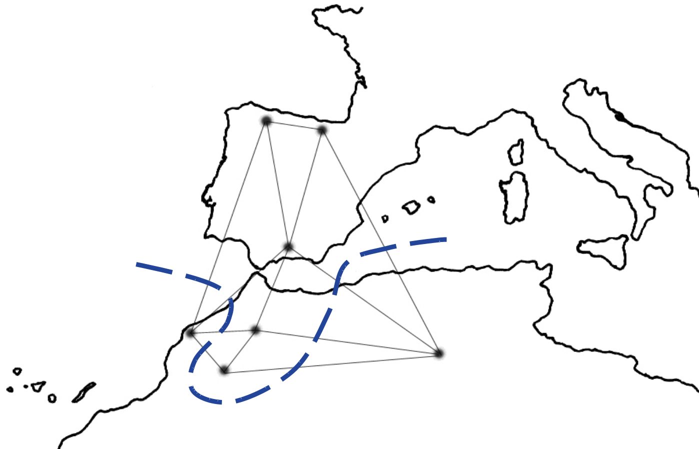

Fig. 5. First genetic boundary (dashed line) recognized on the basis of genetic distances after a Delaunay network analysis among geographical locations of the samples analyzed.

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?

SAŽETAK

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.

9 *Genetic Relationships Among Berbers and South Spaniards Based on CD4 microsatellite/*Alu *haplotypes*

Esther Esteban, Emili González-Pérez, Nourdin Harich, Marc Via, Antoni López-Alomar, Francisco Luna & Pedro Moral.

Annals of Human Biology 2004; 31(2): 202-212

9.1 *Relacions genètiques entre poblacions berbers i andaluses basades en l'anàlisi d'haplotips microsatèl·lit/*Alu *CD4*

Un dels sistemes genètics associats a les insercions *Alu* i que ha mostrat més interès a l'hora de ser caracteritzat per a la reconstrucció poblacional, es basa en l'anàlisi conjunt d'una inserció *Alu* polimòrfica amb un microsatèl·lit (STR) relativament proper a la mateixa i en un grau considerable de desequilibri de lligament. Aquest mètode permet aprofitar les característiques pròpies dels elements *Alu* polimòrfics que ja han estat comentades, en conjunció amb la major variació associada d'aquest altre marcador (STR), per a la reconstrucció d'haplotips d'interès evolutiu i que permeten una anàlisi més fina de la genètica de les poblacions estudiades.

Entre aquests sistemes *Alu*/STR, un dels més utilitzats ha estat el del *locus* CD4, les propietats quantitatives i qualitatives del qual i la seva diversitat haplotípica ha estat caracteritzada en antropologia molecular per clarificar el grau de relació genètica entre diferents poblacions humanes.

El segon treball d'aquesta tesi pretén abordar la variació *Alu*/STR del sistema CD4 en dues poblacions humanes representatives d'ambdues ribes de l'estret de Gibraltar, per tal de disposar d'un segon abordatge parcial a les seves relacions biològiques i possibles afinitats històriques. Així van ser sistemàticament genotipades amb aquest objectiu una població característica de les Alpujarras granadines (sud d'Espanya) i una població berber (Khenifra a l'Atles Mitjà) del Marroc.

El més destacat dels resultats obtinguts ha estat la detecció per primer cop de dues combinacions haplotípiques completament noves, en la població berber estudiada.

Concretament, es van detectar a una freqüència prou significativa les combinacions 75(+) i 80(-) que permeten caracteritzar millor les particularitats d'aquest tipus de poblacions nord-africanes.

Pel que fa al sud de la península Ibèrica, la seva composició genètica per al *locus* CD4 no divergeix significativament d'altres poblacions de la riba nord mediterrània. Únicament, cal destacar la presència en freqüències baixes de les combinacions 90(+) i 130(+), haplotips típics sub-saharians i que podrien estar indicant un cert grau de contacte i flux gènic africà al sud de la península Ibèrica, de baixa magnitud relativa però força continu en el temps.

És interessant, així mateix, destacar la relativament elevada diversitat haplotípica de la població berber de l'Atles Mitjà caracteritzada en aquest projecte, amb 18 haplotips diferents en freqüències polimòrfiques, amb una heterozigositat força destacada dins el nord d'Àfrica (0,85) i amb una elevada freqüència de la combinació 110(-) que ha estat atribuïda a un component ancestral de la regió del nord-oest africà.

Complementàriament, una altra conclusió interessant quan es comparen els resultats amb els disponibles a la bibliografia, fa referència al valor intermedi de flux gènic sud-saharià experimentat per aquesta població berber de l'Atlas marroquí: entre la marcada influència que mostren les poblacions mauritanes del sud i la més relaxada presència de "gens sud-saharians" en berbers del nord de Marroc. Així, aquesta influència sud-sahariana semblaria ser la que millor marca el patró de distribució sud-nord de les poblacions analitzades i les de la bibliografia, amb una freqüència en progressiu descens a mesura que ens allunyem cap al nord del suposat focus poblacional d'influència.

Finalment, i tot i l'efecte de combinacions sud-saharianes presents en poblacions del sud peninsular, el treball no dona suport remarcable a una relació especialment directa entre el nord d'Àfrica i el sud de la península Ibèrica, més enllà del general per a tota aquesta regió de la mediterrània occidental.

9.2 *Informe del director sobre la participació del doctorand*

El Dr. **Pedro Moral Castrillo**, Professor Titular del Departament de Biologia Animal de la Universitat de Barcelona i director de la tesi doctoral **"Variació genètica i evolució d'elements *Alu* recents en poblacions humanes. Inferències biodemo-gràfiques i filogeogràfiques"** presentada pel doctorand **Emili González Pérez**, desitja fer constar que la participació del doctorand en l'elaboració de l'article **"*Genetic Relationships Among Berbers and South Spaniards Based on CD4 microsatellite/Alu haplotypes*"** publicada per la revista *Annals of Human Biology*, ha consistit en les següents tasques principals:

- Processament de mostres: extracció de DNA de les mostres poblacionals utilitzades en el treball (100 %)
- Disseny del treball i selecció de marcadors analitzats, conjuntament amb el Dr. Pedro Moral
- Genotipatge al laboratori de marcadors d'inserció *Alu* i STRs (100 %)
- Elaboració de les bases de dades de resultats (100 %)
- Anàlisi estadística dels resultats, conjuntament amb la Dra. Esther Esteban
- Redacció de l'article (amb la Dra. Esteban i el Dr. Moral)

Complementàriament, cal assenyalar que cap dels coautors d'aquest article ha utilitzat, ni implícitament ni explícitament, els resultats d'aquest treball per a l'elaboració d'una altra tesi doctoral. En conseqüència, aquest treball forma part, en exclusiva, del treball de recerca en el que s'emmarca la tesi del doctorand Emili González Pérez.

Signat: Dr. Pedro Moral Castrillo
Barcelona, 29 de setembre de 2009

ANNALS OF HUMAN BIOLOGY
MARCH–APRIL 2004, VOL. 31, NO. 2, 202–212

Taylor & Francis healthsciences

Genetic relationships among Berbers and South Spaniards based on CD4 microsatellite/Alu haplotypes

E. ESTEBAN†, E. GONZÁLEZ-PÉREZ†, N. HARICH‡, A. LÓPEZ-ALOMAR†, M. VIA†, F. LUNA† and P. MORAL†

†Unitat d'Antropologia, Departament de Biologia Animal, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
‡Départément de Biologie, Faculté des Sciences, Université Chouaïb-Doukkali, El Jadida, Morocco

Received 3 April 2003; in revised form 27 October 2003; accepted 5 November 2003

Summary. *Background*: CD4 STR/Alu haplotype diversity, both for its qualitative and quantitative properties, has been widely used in molecular anthropology to clarify the degree of genetic relationships among human populations.
Aim: CD4 STR/Alu variation was studied in two West Mediterranean samples, Andalusians from La Alpujarra region on the north side of the Gibraltar Strait and Berbers from the south, to ascertain the pattern of affinities between them.
Subjects and methods: Alu and microsatellite alleles were tested in 99 Andalusians from La Alpujarra region (Southeast Spain) and 124 Middle Atlas Berbers (Morocco).
Results: Two new combinations of Alu and STR alleles (75(+) and 80(−)) were found in Berbers. The CD4 STR/Alu haplotype distribution in South Spaniards is similar to that of other Europeans, the only special feature is the slight presence of the 90(+) and 130(+) typical Sub-Saharan haplotypes. The Berber sample is characterized by a high number of different haplotypes (18) with intermediate heterozygosity values (0.846) in comparison with other North African groups, and by a high frequency of the 110(−) combination that has been proposed as representative of an ancient Northwest African population.
Conclusion: A geographical gradient of Sub-Saharan gene contribution has been detected in North Africa. The Middle Atlas Berbers showed an intermediate value in comparison with the high and low values found in Mauritanians and Moroccan Berbers, respectively. The analysis of the CD4 STR/Alu haplotype variation failed to indicate any particular relationship between South Spaniards and North Africans.

1. Introduction

The genetic structure of some Mediterranean populations, largely studied using anthropometric, cultural and genetic data (see, for instance, Jackes *et al.* 1997, Torroni *et al.* 1998, Simoni *et al.* 1999, Richards *et al.* 2000, Comas *et al.* 2000) can be considered the result of a complex pattern of isolation periods and mutual interactions. The populations of the West Mediterranean extreme, represented by the Iberian Peninsula and Morocco, are a good example of this complexity. Different works using classic and DNA markers pointed to a clear differentiation between North Africans and Iberians (Bosch *et al.* 1999, 2000, Comas *et al.* 2000, Brakez *et al.* 2001) while other authors proposed a certain degree of interchange that could date from pre-Neolithic times (Arnaiz-Villena *et al.* 1997). Apart from the possible existence of a common genetic background, the seven centuries of North African domination of the Iberian Peninsula left cultural traces detectable in the present-day

Annals of Human Biology ISSN 0301–4460 print/ISSN 1464–5033 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/03014460310001652275

South Spanish population but the genetic impact of this contact is still a controversial question.

Assessing the genetic impact of seven centuries of coexistence in the past of two groups, probably not very different genetically, is difficult by itself, but it becomes more difficult when the current descendants of these peoples cannot be exactly identified. In this way, Moroccan Berbers can be considered good representatives of the North Africans who invaded the Iberian Peninsula (Hourani 1991). Berbers, the original inhabitants of North Africa, have historically contacted with Phoenicians, Carthaginians, Romans and Arabs who reached North Africa introducing (7th century) and consolidating (11th century) the Islamic religion, rules and language. In the 11th and 13th centuries, the Almoravids and the Almohads—two Berber dynasties—governed large parts of Northwest Africa and Spain. From 711 to 1492 AD South Spain was known as *Al-Andalus*, making reference to the Islamic domination of this region. The longest Muslim permanence in the Iberian Peninsula was recorded in the mountainous and sheer region of La Alpujarra (eastern Andalusia) where a great number of rebel people took refuge after the Moorish expulsion decreed by the Castilian monarchy. The permanence of Moorish people in La Alpujarra is documented until 1571 when, after a local war, they were finally conquered and the Moorish population were replaced by north-west settlers at the end of the 16th century (Trillo 1994). Diverse studies on La Alpujarra population using classic markers (Fernández-Santander *et al.* 1999, Kandil *et al.* 1999) and Alu polymorphisms (González-Pérez *et al.*, 2003) suggested a certain genetic affinity with North Africa, at a higher level than that showed by other North Spanish samples like Basques who remained free of Muslims invaders according to historical information.

Besides the importance of the CD4 gene in the immune response, CD4 short tandem repeat (STR)/Alu haplotype variation has been widely used in molecular anthropology to explain questions so different in magnitude as the origin and dispersion of modern humans (Tishkoff *et al.* 1996) or, more locally, the peopling of concrete regions as the Canary Islands (Flores *et al.* 2001). Concerning Northwest Africa, the study of Flores *et al.* (2000) based on general mixed samples from Morocco and Spain suggested substantial gene flow from Sub-Saharan to Northwest Africa, and from North Africa into the Iberian Peninsula. In this context, the aim of the present study is the analysis of the CD4 STR/Alu variation in two ethnically and geographically well defined samples from Morocco and Spain in order to check previous results. The two samples analysed here are Berber speakers from the Khenifra region in the Middle Atlas mountains (Morocco), and Andalusians from La Alpujarra region in the Sierra Nevada mountains (Southeast Spain). The choice of these two populations goes with the fact that the Khenifra region can be considered as one of the chief regions of Berber culture in Morocco, and in the case of La Alpujarra by the long permanence of Moorish people in this region. The genetic information provided by this polymorphism can be added to previous studies to complete the genetic characterization of these populations and is discussed in relation to the possible biological impact of the Muslim conquest on southern Iberia and the genetic picture of Northwest African populations.

2. Materials and methods

A maximum number of 124 and 99 healthy unrelated individuals from rural areas both of the Moroccan Middle Atlas mountains and La Alpujarra (a region

in the Mediterranean side of the Sierra Nevada mountains, Southeast Spain) were analysed. All individuals had the four grandfathers born in the same region. The sampled Berbers were Tamazight-speakers from the Zayane and Ichkeren confederations (each confederation is a group of neighbour tribes) living in the Khenifra region. The samples were collected by the staff of the Départément de Biologie (El Jadida University) with the help of the Provincial Hospital of Khenifra (Morocco), and by members of the laboratory of Anthropology (University of Barcelona) in collaboration with the Blood Bank Unit of the Torrecárdenas Hospital (Almería, Spain). Appropriate informed consent was obtained from all participating individuals.

Genomic DNA was extracted from blood by standard phenol–chloroform methods. PCR amplification conditions for the Alu and the microsatellite polymorphisms and allele nomenclature were according to Tishkoff *et al.* (1996) and Flores *et al.* (2000). In the microsatellite genotyping, STR alleles were identified using specific ladders of known CD4 microsatellite alleles spanning the entire range of STR sizes.

Allele frequencies were estimated by direct gene counting. Exact tests were used to check for Hardy–Weinberg equilibrium and to compare frequency distribution among samples. Locus heterozygosity was estimated according to Nei's formula (Nei 1987). Maximum likelihood haplotype frequencies were computed using the EM algorithm. Global relationships were estimated by means of Reynolds' genetic distances and by a principal component analysis (PCA) based on the correlation matrix of the CD4 STR/Alu haplotype frequencies by using the SPSS package. Calculations were performed using the GENEPOP v.3.3 (Raymond and Rousset 1995), ARLEQUIN (Schneider *et al.* 2000) and PHYLIP (Felsenstein 1989) statistical packages. Genetic admixture was tested by means of the ADMIX1.0 program (Bertorelle and Excoffier 1998) providing three different estimators: a least-squared estimator (mR), a maximum-likelihood estimator (mC) and an estimator based on mean coalescence times (mY).

For comparative purposes, data from several historically and/or geographically related populations were selected from the literature. The Iberian Peninsula was represented by Basques from North Spain (S. Tishkoff and P. Moral, personal communication), Canary Islanders (Flores *et al.* 2001) and a miscellaneous sample of Spaniards (Flores *et al.* 2000). Regarding Northwest Africa, four groups reported by Flores *et al.* (2000) were used: (a) Arab-speaking Moroccans from Melilla and Agadir, (b) Berbers from the Moroccan localities of Melilla, Central Atlas, Rif mountains and Agadir, (c) West Saharans (Berber speaking) from Tindouf (Morocco), and (d) Mauritanians from Nouatchok and Nouadhibou. Finally, a mixed sample of Europeans (Tishkoff *et al.* 1996) and Sub-Saharan Senufo from the Guinea Gulf (Flores *et al.* 2000) were included as external references.

3. Results

CD4 microsatellite, Alu and haplotype observed frequencies are shown in table 1. Both polymorphism (STR and Alu) distributions are in Hardy–Weinberg equilibrium in the two populations. It is interesting to note the occurrence of a low-range allele of 75 bp unambiguously associated with the Alu insertion (Alu+) in one Berber chromosome (individual genotype: 75/115 for the STR, and +/+ for the Alu). This 75 bp STR allele is common in most hominoid species (Tishkoff *et al.* 1996) in combination with the Alu insertion (Alu+) since Alu deletion (Alu−) is absent in

Table 1. CD4 microsatellite and Alu allele frequencies. CD4 STR/Alu haplotypes in Andalusians from La Alpujarra and Moroccan Berbers.

		Frequency				Frequency	
Locus	Alleles	Andalusians	Berbers	Locus	Alleles	Andalusians	Berbers
CD4 Alu	Alu+	0.7020	0.6089	CD4 STR	75	–	0.0042
	Alu−	0.2980	0.3911		80	–	0.0125
		$n=99$	$n=124$		85	0.3239	0.3792
		$H=0.418$	$H=0.478$		90	0.3125	0.2167
					95	–	0.0167
					100	–	0.0333
					105	0.0057	0.0167
					110	0.3011	0.2000
					115	0.0341	0.0917
					120	–	0.0208
					125	0.0057	0.0083
					130	0.0170	–
						$n=88$	$n=120$
						$H=0.705$	$H=0.759$

	Frequency ± SE				Frequency ± SE		
Haplotypes	Andalusians $n=87$	Berbers $n=118$	Ranges	Haplotypes	Andalusians	Berbers	Ranges
75+	–	0.0042 ± 0.0037	–	80−	–	0.0127 ± 0.0072	–
85+	0.3276 ± 0.0346	0.2901 ± 0.0316	0.171 WS to 0.553 AM	85−	–	0.0955 ± 0.0219	0 SP to 0.115 SS
90+	0.0173 ± 0.0098	0.0056 ± 0.0051	0 AM to 0.083 WSS	90−	0.2987 ± 0.0367	0.2147 ± 0.0265	0.007 OC to 0.327 CI
100+	–	0.0172 ± 0.0095	0 SP to 0.165 SS	95−	–	0.0169 ± 0.0074	0 SP to 0.056 WSS
105+	0.0057 ± 0.0055	0.0062 ± 0.0058	0 ME to 0.037 WSS	100−	–	0.0166 ± 0.0097	0 SP to 0.047 MA
110+	0.2872 ± 0.0368	0.1516 ± 0.0238	0.125 WS to 0.421 AM	105−	–	0.0107 ± 0.0079	0 SP to 0.019 SS
115+	0.0345 ± 0.0135	0.0573 ± 0.0179	0.004 AM to 0.188 WS	110−	0.0058 ± 0.0056	0.0433 ± 0.0188	0 SP to 0.047 MA
120+	–	0.0101 ± 0.0073	0 SP to 0.107 SS	115−	–	0.0275 ± 0.0120	0 SP to 0.027 SS
125+	0.0057 ± 0.0056	0.0085 ± 0.0071	0 SP to 0.060 SS	120−	–	0.0110 ± 0.0068	0 SP to 0.016 WS
130+	0.0172 ± 0.0101	–	0 SP to 0.053 CI				

n = Sample size, H = heterozygosity.
Population abbreviations in ranges correspond to: AM: Amerindians, CI: Canary Islanders, MA: Mauritanians, ME: Middle Easterns, OC: Oceanians, SP: several populations, SS: Sub-Saharans, WS: West Saharans, WSS: West Sub-Saharans.

non-human primates. Previously, the presence of the 75 bp allele was only described in the Canarian Island of Lanzarote (Flores *et al.* 2001) but associated with the Alu deletion. Also, a new combination of a 80 bp allele and the Alu deletion (80(−)) was detected as double homozygote in a subject of the Berber sample.

Although no significant differences were found between the two populations examined for each locus, the haplotype distribution was clearly distinct as evidenced the exact test of sample differentiation ($p = 0.0045 \pm 0.0018$). Middle Atlas Berbers show 2 times greater different haplotypes (18) than La Alpujarra (nine) and consistently the heterozygosity values in Berbers (0.846 ± 0.015) are higher than in the other sample (0.795 ± 0.015). Haplotype frequencies in La Alpujarra fit well into the variation described for European populations while the relatively high values (see ranges of table 1) of the 85(−) and the 110(−) combinations in Berbers are consistent with Sub-Saharan and Northwest African influences, respectively.

The common worldwide population pattern (see ranges in table 1) shows that, excluding the haplotypes showing frequencies lower than 1%, most samples exhibit higher number of haplotypes with the Alu insertion than those carrying the Alu deletion in accordance with the fact that the ancestral state is the full-length Alu. The Alu(+)/Alu(−) haplotype rates are 10/3 in Sub-Saharans, 8/5 in West-Saharans, 6/4 in Moroccans, 4/2 in Spanish, 5/1 in La Alpujarra and Europeans. However, this is not the case for the Middle Atlas Berbers and Mauritanians that show, respectively, an inverse pattern of 5/9 and 6/7 Alu(+) vs Alu(−) haplotypes. Relative weighted mean frequencies of the STR alleles standardized for Alu(+) and Alu(−) chromosomes are given in table 2. The distribution of STR alleles on Alu(+) chromosomes shows similar heterozygosity values in all populations (ranges from 0.56 of Andalusia to 0.84 of Sub-Saharan Africa) while the STR distribution on Alu(−) chromosomes is remarkably different. The STR pattern in Alu(−) chromosomes clearly distinguishes two different groups of populations, one formed by Iberians and some North Africans showing extremely high frequencies of the 90 bp allele (the STR allele on which the Alu deletion appeared) and subsequently low heterozygosities (from 0.00 in Basques to 0.33 in Berbers), and a group formed by Sub-Saharans and several North African populations that showed great STR allele variation and high diversity values (from 0.59 of West Saharans to 0.78 of Sub-Saharan Africa). The Middle Atlas Berbers (heterozygosity of 0.70) come in this last group.

The high diversity values found in some of the North African groups mentioned above and the presence of characteristic Sub-Saharan haplotypes (115(+) and 85(−)) in these populations suggest a considerable degree of Sub-Saharan contribution that could be measured through frequency-based estimators of admixture, even though the results should be taken with caution because they are based on a single locus. In our case, we have measured the Sub-Saharan contribution to North Africans taking Europeans (Tishkoff *et al.* 1996) and Senufo (Flores *et al.* 2000) as parental populations, but we are aware that a mixed European sample is not the best representation of the ancestral North African parental population and, probably, will magnify the Sub-Saharan contribution. The three admixture estimators (mR, mC and mY) yielded very similar values, with those from mR being the least extreme. Admixture values (mR) were, in decreasing order, 63.93% (SD = 0.14) in West Saharans, 63.09% (SD = 0.15) in Mauritanians, 30.04% (SD = 0.07) in Middle Atlas Berbers, 13.84% (SD = 0.08) in the Moroccan Berbers, and finally, 11.49% (SD = 0.14) in Arab-speaking Moroccans. These Sub-Saharan contributions are simply suggestive of a considerable south to north gene flow in Africa.

Table 2. Relative weighted frequencies of the different STR alleles standardized separately for Alu(+) or Alu(−) chromosomes in different populations.

Population	STR allele frequencies													
	75	80	85	90	95	100	105	110	115	120	125	130	135	*H*
STR allele frequencies in Alu(+) chromosomes														
Spain*	–	–	0.4779	0.0191	–	0.0088	0.0088	0.3930	0.0660	0.0088	0.0088	0.0088	–	0.6120
Andalusia†	–	–	0.4712	0.0249	–	–	0.0082	0.4131	0.0497	–	0.0082	0.0247	–	0.6035
Basque Country‡	–	–	0.4875	0.0250	–	–	–	0.4375	0.0375	0.0125	–	–	–	0.5688
Arab-speaking Moroccans*	–	–	0.3583	0.0190	–	–	0.0570	0.3962	0.1505	0.0190	–	–	–	0.6878
Moroccan Berbers*	–	–	0.4021	0.0577	–	0.0410	0.0076	0.3763	0.0910	–	0.0167	–	0.0076	0.6830
Middle Atlas Berbers†	0.0076	–	0.5268	0.0102	–	0.0312	0.0112	0.2753	0.1040	0.0183	0.0154	–	–	0.6341
W. Sahara*	–	–	0.2490	0.1136	–	0.0232	–	0.1819	0.2737	0.0684	0.0451	0.0451	–	0.8078
Mauritania*	–	–	0.2480	0.0615	–	0.0904	–	0.2482	0.3211	–	0.0308	0.0308	–	0.7609
Sub-Saharan Africa*	–	0.0010	0.2030	0.0280	0.0400	0.0900	0.0300	0.2100	0.2000	0.1400	0.0300	0.0300	0.0040	0.8400
STR allele frequencies in Alu(−) chromosomes														
Spain		–	0.0189	0.8647	–	–	–	0.0818	0.0157	0.0189	–	–		0.2446
Andalusia		–	–	0.9809	–	–	–	0.0191	–	–	–	–		0.0375
Basque Country		–	–	1.0000	–	–	–	–	–	–	–	–		0.0000
Arab-speaking Moroccans		–	0.0326	0.9022	–	0.0326	–	0.0326	–	–	–	–		0.1828
Moroccan Berbers		–	0.0469	0.8095	–	0.0322	–	0.0322	0.0645	–	0.0147	–		0.3361
Middle Atlas Berbers		0.0283	0.2127	0.4784	0.0376	0.0370	0.0238	0.0964	0.0613	0.0245	–	–		0.7081
W. Sahara		–	0.0990	0.6007	–	–	–	0.0990	0.1502	0.0511	–	–		0.5944
Mauritania		–	0.2021	0.4042	0.0667	0.0979	–	0.0979	0.0979	0.0333	–	–		0.7615
Sub-Saharan Africa		–	0.2640	0.2580	0.0680	0.0410	0.0080	–	0.2710	0.0450	0.038	0.007		0.7800

H: heterozygosity calculated using $1 - \Sigma f^2$; *f* is the standardized frequency shown in the table for each STR allele.
Populations references: * Flores *et al.* 2000; †this study; ‡Tishkoff and Moral, personal communication.

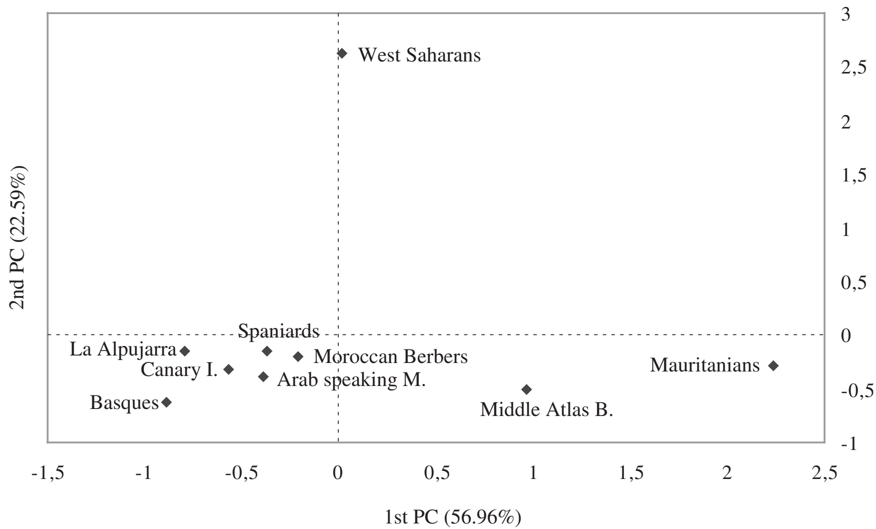

Figure 1. Principal components plot based on CD4 STR/Alu haplotype frequencies in North African and Iberian populations.

As for Northwest African and Iberian CD4 STR/Alu distributions, pairwise comparisons showed significant differences of Middle Atlas Berbers from Basques, Canary Islanders and West Saharans on the one hand, and between La Alpujarra and West Saharans and Mauritanians on the other. Genetic differentiation among populations has been analysed by computing Reynolds genetic distances and depicted by a principal component (PC) plot (figure 1). The first two axes account for 73.9% of the genetic variance observed. The first PC clearly distinguishes Mauritanians and, to a lesser degree, Middle Atlas Berbers, from the rest of populations while the second PC separates West Saharans. The typical Sub-Saharan 115(–) and 115(+) haplotypes along with the 110(–) haplotype reported as characteristic of North Africans are highly correlated (more than 85%) with the positive values of first PC. The haplotypes 120(+), 125(+) and 90(+), which show high values in West Saharans in comparison with the frequencies described in other Africans, are best correlated (more than 75%) with the differentiation of this population. In general terms, the pattern of Reynolds distances matches the population picture of the PCA plot, but a detailed analysis of the distances matrix demonstrates some population relationships that are disguised by the heterogeneous cluster of figure 1. There is not a clear and consistent pattern of differentiation among Moroccans and the Iberian Peninsula since the genetic distances into each region (distance ranges from 0.0017 to 0.0100 in Spaniards, and 0.0055 to 0.0241 in Moroccans) are sometimes higher than those between regions (from 0.0045 to 0.0492) but, if we consider only the geographically well defined samples, the pattern of genetic distances among North Africa and Spain becomes consistent: Middle Atlas Berbers showed the same genetic distance to Mauritania (0.0239) than to La Alpujarra (0.0249), and 2 times higher distance to North Spanish Basques (0.0429). This pattern of lower distances to South Spain in comparison with the Basque sample is also shown by all the other North African groups. However, the genetic differentiation among these nine Mediterranean groups

based on the CD4 STR/Alu haplotype frequencies is weak (global $F_{ST} = 0.8\%$, $p = 0.008$).

4. Discussion

The obtained CD4 STR/Alu data on two anthropologically well-defined samples complete the gene information available on South Spaniards and Moroccan Middle Atlas Berbers and allow us to discuss some questions about the genetic background of these populations. Although the analysis of only two polymorphisms in a single gene demands caution, the fact that some CD4 haplotypes were characteristic of Sub-Saharan populations are especially useful for detection of admixture into the Iberian Peninsula and North Africa.

La Alpujarra population in southern Spain appears clearly similar to other Iberians and Europeans, only the frequencies of the 90(+) and 130(+) typical Sub-Saharan haplotypes (1.7% in both cases) are slightly higher than in other Spanish samples (1.3% and 0.6%). The remarkable differentiation between Middle Atlas Berbers and La Alpujarra, reflected in the PCA plot, contrasts with data from other estimates of North African contribution into the Iberian Peninsula. Estimates of North African gene flow in the general Spanish population based on the CD4 locus (19.5%, Flores *et al.* 2000) and other different markers such as STR (31%, Bosch *et al.* 2000) and mtDNA (18%, Plaza *et al.* 2003) indicated a relevant North African contribution while classical data (Harich *et al.* 2002) failed to find any important affinity among North Africans and Iberians.

As for North African populations, CD4 genetic data confirms a remarkably higher diversity in this region than in the Northern coast of the Mediterranean (genetic distance average: 0.0253 vs 0.0053). Our results on the CD4 distribution in the Middle Atlas Berber sample reveals several distinctive features of North African populations. The relatively high gene diversity is consistent with the high number of different haplotypes (18) observed in the Middle Atlas Berbers and their relatively high heterozygosity (0.846 ± 0.015), only slightly lower than in other southernmost North African samples (West Sahara and Mauritania, $H = 0.889$).

Concerning the pattern of the CD4-STR allele distribution among Alu +/− chromosomes, the highest Alu(−) diversity has been reported for Sub-Saharan Africans while the Alu deletion is almost exclusively associated to the STR 90 allele (Tishkoff *et al.* 1996) in non-Africans with the exception of North-Saharan African populations (Flores *et al.* 2000). In this context, the Alu(−) distribution of STR alleles observed in the Middle Atlas Berbers is more similar to Mauritanians than to other Berber and Arab Moroccan groups, suggesting a remarkable degree of Sub-Saharan genetic influence that was estimated as 30% from our CD4 data. This value is comparable with several admixture values provided by other authors on the basis of autosomal restriction fragment length polymorphism (RFLP) (25.5%, Fernandez-Santander *et al.* 2002) and mtDNA (26.1%, Gonzalez *et al.* 2003; 25.9%, Plaza *et al.* 2003) although values from the Y chromosome seem to be considerably lower (8%, Bosch *et al.* 1999).

Finally, it is worth noting the noticeable presence (4.3%) of the 110(−) haplotype in the Middle Atlas Berber population (West Mediterranean variation: 0.5–2.6% in the Iberian Peninsula, 1.1–4.7% in North Africa) confirming that conspicuous frequencies of this haplotype can be distinctive genetic characteristics of North Africa (Flores *et al.* 2000). This result, including the trail of this haplotype in La Alpujarra along with other Iberian samples, might be consistent with the

hypothesis about the existence of a common genetic background in Western Mediterranean generated by Upper Palaeolithic population expansions (Torroni *et al*. 1998, Richards *et al*. 2000).

To sum up, this study of the CD4 STR/Alu variation demonstrates the existence of clearly different distributions between southern Iberia and North Africa through the analysis of two strictly geographical and ethnical defined population samples (La Alpujarra and Middle Atlas Moroccan Berbers), more important than those reported by other studies from general population samples. Besides, in comparison with previous studies the data on the Middle Atlas Berber populations confirm the presence of the 110(−) haplotype in Northwest Africa, the distinctive features of the Alu(−) haplotype distribution in this region, and are consistent with remarkable Sub-Saharan genetic influence in these populations that could have been an important factor for the relative current differentiation between the two shores of the Mediterranean region.

Acknowledgements

This work has been supported in part by the Agencia Española de Cooperación Internacional, 1998SGR00129 and 2001SGR00089 grants, and by Comissionat per a Universitats i Recerca de la Generalitat de Catalunya grants (1998FI00664, 2001FI00177, 2002FI 00516).

References

Arnaiz-Villena, A., Martínez-Laso, J., Gómez-Casado, E., Díaz-Campos, N., Santos, P., Martinho, A., and Breda-Coimbra, E., 1997, Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes. *Immunogenetics*, **47,** 37–43.

Bertorelle, G., and Excoffier, L., 1998, Inferring admixture proportions from molecular data. *Molecular Biology and Evolution*, **15,** 1298–1311.

Bosch, E., Clafell, F., Santos, F. R., Perez-Lezaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., and Bertranpetit, J., 1999, STR variation is deeply structured by genetic background on the human Y chromosome. *American Journal of Human Genetics*, **65,** 1623–1638.

Bosch, E., Calafell, F., Perez-Lezaun, A., Clarimón, J., Comas, D., Mateu, E., Martínez-Arias, R., Morera, B., Brakez, A., Akhayat, O., Sefiani, A., Hariti, G., Cambon-Thomsen, A., and Bertranpetit, J., 2000, Genetic structure of north-west Africa revealed by STR analysis. *European Journal of Human Genetics*, **8,** 360–366.

Brakez, A., Bosch, E., Isabel, H., Akhayat, O., Comas, D., Bertranpetit, J., and Calafell, F., 2001, Human mitochondrial DNA sequence variation in the Moroccan populations of the Souss area. *Annals of Human Biology*, **28,** 295–307.

Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefranc, G., Stoneking, M., Batzer, M. A., Bertranpetit, J., and Sajantila, A., 2000, Alu insertion polymorphisms in NW Africa and the Iberian Peninsular: evidence for a strong genetic boundary through the Gibraltar Straits. *Human Genetics*, **107,** 312–319.

Felsenstein, J., 1989, PHYLIP: Phylogeny Inference Package (version 3.2). *Cladistics*, **5,** 164–166.

Fernández-Santander, A., Kandil, M., Luna, F., Esteban, E., Jiménez, F., Zaoui, D., and Moral, P., 1999, Genetic relationships between Southeastern Spain and Morocco: new data on ABO, RH, MNSs, and DUFFY polymorphisms. *American Journal of Human Biology*, **11,** 745–752.

Fernandez-Santander, A., Kandil, M., Luna, F., and Moral, P., 2002, Twenty nuclear DNA polymorphisms in a Moroccan population: a comparison with seven other human populations. *Human Biology*, **74,** 695–706.

Flores, C., Maca-Meyer, N., González, A. M., and Cabrera, V. M., 2000, Northwest African distribution of the CD4/Alu microsatellite haplotypes. *Annals of Human Genetics*, **64,** 321–327.

Flores, C., Maca-Meyer, N., Pérez, J. A., and Cabrera, V. M., 2001, The peopling of the Canary Islands: A CD4/Alu Microsatellite haplotype perspective. *Human Immunology*, **62,** 949–953.

Gonzalez, A. M., Brehm, A., Perez, J. A., Maca-Meyer, N., Flores, C., and Cabrera, V. M., 2003, Mitochondrial DNA affinities at the Atlantic fringe of Europe. *American Journal of Physical Anthropology*, **120,** 391–404.

González-Pérez, E., Via, M., Esteban, E., Lopez-Alomar, A., Mazieres, S., Harich, N., Kandil, M., Dugoujon, J. M., and Moral, P., 2003, Alu insertions in the Iberian Peninsula and NW Africa: genetic boundaries or melting pot? *Collegium Antropologicum*, **27**, 491–500.

Harich, N., Esteban, E., Chafik, A., López-Alomar, A., Vona, G., and Moral, P., 2002, Classical polymorphisms in Berbers from Moyen Atlas (Morocco): genetics, geography, and historical evidence in the Mediterranean peoples. *Annals of Human Biology*, **39,** 473–487.

Hourani, A. H., 1991, *A History of the Arab Peoples* (Cambridge, MA: Belknap Press of Harvard University).

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

Kandil, M., Moral, P., Esteban, E., Autori, L., Mameli, G. E., Zaoui, D., Calo, C., Luna, F., Vacca, L., and Vona, G., 1999, Red cell enzyme polymorphisms in Moroccans and Southeastern Spaniards: new data for the genetic history of the Western Mediterranean. *Human Biology*, **71,** 791–802.

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

Plaza, S., Calafell, F., Helal, A., Bouzerna, N., Lefranc, G., Bertranpetit, J., and Comas, D., 2003, Joining the Pillars of Hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean. *Annals of Human Genetics*, **67,** 312–328.

Raymond, M., and Rousset, F., 1995, GENEPOP: population software for exact tests and ecumenicism. *Journal of Heredity*, **86,** 248–249.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., and Stefanescu, G., 2000, Tracing European founder lineages in the Near Eastern mtDNA. *American Journal of Human Genetics*, **67,** 1251–1276.

Schneider, S., Roessli, D., and Excoffier, L., 2000, Arlequin (ver. 2000). A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Simoni, L., Gueresi, P., Pettener, D., and Barbujani, G., 1999, Patterns of gene flow inferred from genetic distances in the Mediterranean region. *Human Biology*, **71,** 399–415.

Tishkoff, S. A., Dietzsch, E., Speed, W., Pakstis, A. J., Kidd, J. R., Cheung, K., Bonné-Tamir, B., Santachiara-Benerecetti, A. S., Moral, P., Krings, M., Pääbo, S., Watson, E., Risch, N., Jenkins, T., and Kidd, K. K., 1996, Global patterns of linkage disequilibrium at the CD4 locus and modern human populations. *Science*, **271,** 1380–1387.

Torroni, A., Bandelt, H. J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonné-Tamir, B., and Scozzari, R., 1998, mtDNA analysis reveals a major late Paleolithic expansion from Southwestern to Northeastern Europe. *American Journal of Human Genetics*, **62,** 1137–1152.

Trillo, C., 1994, La Alpujarra antes y después de la conquista castellana (Granada: Universidad de Granada).

Address for correspondence: Dr Pedro Moral, Unitat d'Antropologia, Departament de Biologia Animal, Facultat de Biologia, Universitat de Barcelona, Avda Diagonal 645, 08028 Barcelona, Spain. email: pmoral@ub.edu

Zusammenfassung. *Hintergrund:* In der molekularen Anthropologie ist die CD4 STR/Alu–Haplotypenvariabilität sowohl in qualitativer als auch in quantitativer Hinsicht vielfach benutzt worden, um den Grad der genetischen Beziehungen zwischen menschlichen Populationen zu klären.

Zielstellung: Die CD4 STR/Alu–Variabilität in zwei westmediterranen Stichproben (Andalusier aus der Region La Alpujarra auf der Nordseite der Straße von Gibraltar und Berber aus dem Süden) wird verwendet, um die Ähnlichkeit zwischen ihnen festzustellen.

Material und Methode: Bei 99 Andalusiern aus der Region La Alpujarra (Südostspanien) und 124 Berbern aus dem mittleren Atlas wurden Alu- und Mikrosatellitenallele untersucht.

Ergebnisse: Bei Berbern wurden zwei neue Kombinationen von Alu- und STR-Allelen (75(+) und 80(−)) gefunden. Die CD4 STR/Alu–Verteilung bei Südspaniern ähnelt derjenigen anderer Europäer; das einzige spezifische Merkmal ist die geringe Rate der typischen Subsahara-Haplotypen 90(+) und 130(+). Die Stichprobe der Berber wird durch eine hohe Zahl verschiedener Haplotypen (18) mit intermediären Heterozygotie-Werten (0.846) im Vergleich mit anderen nordafrikanischen Gruppen und eine hohe Frequenz der 110(−)-Kombination charakterisiert, welche als repräsentativ für eine alte nordwestafrikanische Population gilt.

Schlussfolgerungen: In Nordafrika wurde ein geographischer Gradient der Subsahara-Gen-Verteilung festgestellt. Im Vergleich mit den hohen und niedrigen Werten bei mauretanischen beziehungsweise marokkanischen Berbern weisen die Berber aus dem mittleren Atlas intermediäre Werte auf. Bei der Analyse der CD4 STR/Alu–Haplotypen-Variabilität konnte keine bestimmte Beziehung zwischen Südspaniern und Nordafrikanern nachgewiesen werden.

Résumé. *Arrière-plan:* La diversité haplotypique CD4 STR/Alu a à la fois pour ses propriétés qualitatives et quantitatives, été largement utilisée en anthropologie moléculaire pour clarifier le degré de relation génétique des populations humaines.
But: La variation CD4 STR/Alu est utilisée dans deux échantillons ouest-méditerranéens : andalous de la région de la Alpujarra (sud-est de l'Espagne) au nord du détroit de Gibraltar et berbères au sud du même détroit, pour étudier leurs affinités.
Sujets et méthodes: On a testé les allèles Alu et microsatellites chez 99 andalous de la région de la Alpujarra (sud-est de l'Espagne) et chez 124 berbères du Moyen-Atlas.
Résultats: Deux nouvelles combinaisons d'allèles Alu et STR (75(+) et 80(−) ont été trouvées chez les berbères. La distribution de l'haplotype CD4 STR/Alu chez les espagnols du sud est analogue à celle des autres européens, le seul élément particulier étant la faible présence des haplotypes 90(+) et 130(+) typiques de l'Afrique sub-saharienne. L'échantillon berbère est caractérisé par un nombre élevé (18) d'haplotypes distincts avec des valeurs d'hétérozygosité intermédiaires (0,846) par rapport à d'autres groupes nord-africains et par une haute fréquence de la combinaison 110(−) qui est supposée représenter une population nord-africaine ancienne.
Conclusion: Un gradient géographique de contribution de gènes d'Afrique sub-saharienne a été détecté en Afrique du Nord, les berbères du Moyen-Atlas montrant des valeurs intermédiaires par rapport aux valeurs hautes et basses trouvées respectivement chez les mauritaniens et chez les berbères du Maroc. L'analyse de la variation de l'haplotype CD4 STR/Alu ne révèle pas de relation particulière entre espagnols du sud et nord-africains.

Resumen. *Antecedentes:* La diversidad del haplotipo CD4 STR/Alu, por sus propiedades tanto cualitativas como cuantitativas, se ha utilizado ampliamente en antropología molecular para clarificar el grado de relación genética entre poblaciones humanas.
Objetivo: la variación CD4 STR/Alu se ha utilizado en 2 muestras del Mediterráneo occidental, Andaluces de la región de La Alpujarra, en la zona norte del Estrecho de Gibraltar y Beréberes del Sur, para dilucidar el patrón de afinidades entre ellos.
Sujetos y Métodos: Se han probado los alelos Alu y microsatélite en 99 Andaluces de la región de La Alpujarra (SE de España) y en 124 Beréberes del Atlas Central.
Resultados: Se han encontrado 2 nuevas combinaciones de alelos Alu y STR (75(+) y 80(−)) en Beréberes. La distribución del haplotipo CD4 STR/Alu en los españoles del Sur es similar a la de otros europeos, siendo la única característica especial la débil presencia de 90(+) y 130(+), típica de los haplotipos subsaharianos. La muestra Bereber se caracteriza por un elevado número de haplotipos diferentes (18) con valores medios de heterozigosidad (0,846) en comparación con otros grupos africanos, y por una elevada frecuencia de la combinación 110(−), que ha sido propuesta como representativa de una antigua población del Noroeste africano.
Conclusión: Se ha detectado un gradiente geográfico de contribución génica sub-sahariana en el norte de África, los Beréberes del Atlas Central mostraban un valor intermedio en comparación con los valores altos y bajos encontrados en mauritanos y Beréberes marroquíes, respectivamente. El análisis de la variación del haplotipo CD4 STR/Alu no permite indicar ninguna relación particular entre los españoles del Sur y los Norteafricanos.

10 *Population Relationships in the Mediterranean Revealed by Autosomal Genetic Data (*Alu *and* Alu*/STR Compound Systems)*

Emili González-Pérez, Esther Esteban, Marc Via, Magdalena Gaya-Vidal, Georgios Athanasiadis, Jean-Michel Dugoujon, Francisco Luna, Mª Soledad Mesa, Vicente Fuster, Mostafa Kandil, Nourdin Harich, Nisrine Bissar-Tadmouri, Angela Saeta & Pedro Moral.

American Journal of Physical Anthropology 2009 (in press)

10.1 *Relacions entre poblacions mediterrànies revelades per dades genètiques autosòmiques (elements* Alu *i sistemes compostos* Alu*/STR)*

Aquest estudi consolida l'estratègia aplicada en els dos treballs previs per a tot l'àmbit continental circummediterrani i a partir d'una bateria molt completa de marcadors genètics de diferent naturalesa i que han demostrat la seva utilitat en la caracterització de les relacions evolutives entre poblacions humanes, així com la seva capacitat per discernir processos històrics i biodemogràfics recents i els processos particulars de poblament humà per a la regió mediterrània.

El treball es basa en la determinació de la variació de 18 elements *Alu* autosòmics així com l'element específic del cromosoma Y (YAP) en un total de 17 poblacions continentals del voltant del Mediterrani. A més a més, s'ha utilitzat la variació en tres sistemes compostos *Alu*/STR que inclouen tres dels elements *Alu* estudiats en un primer moment (els dels *loci* CD4, FXIIIB i DM).

L'objectiu principal d'aquest enfocament era analitzar noves dades sobre les relacions existents entre poblacions humanes d'aquesta regió geogràfica i aportar noves perspectives, a partir de marcadors autosòmics i de sistemes genètics de naturalesa diferenciada, per abordar els processos antics de poblament de la regió i les influències històriques recents que hagin configurat la composició biològica actual d'aquestes poblacions. Els resultats pretenen complementar les dades inicials aportades per altres tipus

d'estudis previs, basats principalment en marcadors clàssics i en marcadors genètics de tipus uniparental (DNA mitocondrial i cromosoma Y).

Entre les primeres observacions cal destacar que els patrons de variació que aporten uns i altres conjunts de marcadors són diferents entre ells, una prova a favor de l'evidència que la naturalesa i taxa de mutació de cada tipus de polimorfisme genètic resulta primordial a l'hora d'interpretar correctament les observacions i les inferències realitzades a partir dels mateixos. En aquest sentit, mentre que la variació genètica de la bateria de marcadors *Alu* bial·lèlics pot ser interpretada com el fruit d'una considerable heterogeneïtat global i de base entre poblacions mediterrànies, la utilització d'haplotips combinats *Alu*/STR semblen detectar amb més detall el grau de diferenciació genètica entre les dues ribes mediterrànies (nord i sud).

Sembla ser que aquesta estructuració entre les dues ribes estaria determinada en primer lloc per l'efecte del flux genètic sud-saharià cap al nord d'Àfrica i des d'aquí –i en menor grau– cap a la riba nord mediterrània. Així, s'ha pogut detectar un grau de contribució sud-sahariana al nord d'Àfica que s'apropa al 13 % quan analitzem la bateria de marcadors *Alu,* però que pot arribar a multiplicar-se per tres quan estudiem els tres *loci* independents de sistemes combinats *Alu*/STR (CD4, FXIIIB i DM). Els resultats globals, confirmen globalment la permeabilitat del Sàhara a les migracions humanes així com l'existència d'intercanvis i moviments humans trans-mediterranis des d'antic.

També entre les conclusions destacades de l'estudi destaca la identificació de dues combinacions específiques *Alu*/STR –la CD4 110(-) i la DM 107(-)–, l'origen de les quals ha estat datat en temps paleolítics (entre els 36 000 i els 48 000 anys) i que semblen ser remanents d'un component ancestral comú dels pobles humans de l'àmbit mediterrani. Així, aquestes combinacions –filogeogràficament exclusives d'aquest àmbit– es troben en freqüències destacades en totes les poblacions berbers nord-africanes, a la península Ibèrica, al sud de França, al Mediterrani oriental (Grècia i Turquia) i en la majoria de les grans illes mediterrànies.

El treball demostra, en conjunt, la capacitat d'aquest nou abordatge de diferents sistemes i marcadors autosòmics per detectar adequadament relacions poblacionals a petita escala i en detall, amb patrons específics que reconeixen orígens antics comuns i influències biològiques compartides per a totes les poblacions circummediterrànies.

10.2 Informe del director sobre la participació del doctorand

El Dr. **Pedro Moral Castrillo**, Professor Titular del Departament de Biologia Animal de la Universitat de Barcelona i director de la tesi doctoral **"Variació genètica i evolució d'elements *Alu* recents en poblacions humanes. Inferències biodemo-gràfiques i filogeogràfiques"** presentada pel doctorand **Emili González Pérez**, desitja fer constar que la participació del doctorand en l'elaboració de l'article **"Population Relationships in the Mediterranean Revealed by Autosomal Genetic Data (*Alu* and *Alu*/STR Compound Systems)"** publicada per la revista *American Journal of Physical Anthropology*, ha consistit en les següents tasques principals:

- Processament de mostres: extracció de DNA de les mostres poblacionals utilitzades en el treball (100 %)
- Disseny del treball i selecció de marcadors analitzats, conjuntament amb el Dr. Pedro Moral
- Genotipatge d'insercions *Alu* i STRs / Seqüenciacions (100%)
- Elaboració de les bases de dades de resultats (100 %)
- Anàlisi estadística dels resultats, conjuntament amb la Dra. Esther Esteban
- Redacció de l'article (amb la Dra. Esteban i el Dr. Moral)

Complementàriament, cal assenyalar que cap dels coautors d'aquest article ha utilitzat, ni implícitament ni explícitament, els resultats d'aquest treball per a l'elaboració d'una altra tesi doctoral. En conseqüència, aquest treball forma part, en exclusiva, del treball de recerca en el que s'emmarca la tesi del doctorand Emili González Pérez.

Signat: Dr. Pedro Moral Castrillo
Barcelona, 29 de setembre de 2009

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 000:000–000 (2010)

Population Relationships in the Mediterranean Revealed by Autosomal Genetic Data (*Alu* and *Alu*/STR Compound Systems)

Emili González-Pérez,[1] Esther Esteban,[1] Marc Via,[1,2,3] Magdalena Gayà-Vidal,[1] Georgios Athanasiadis,[1] Jean Michel Dugoujon,[4] Francisco Luna,[5] Maria Soledad Mesa,[5] Vicente Fuster,[5] Mostafa Kandil,[6] Nourdin Harich,[6] Nisrine Bissar-Tadmouri,[7] Angela Saetta,[8] and Pedro Moral[1]*

[1]*Secció d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Barcelona, Spain*
[2]*Department of Medicine, University of California, San Francisco, CA*
[3]*Institute for Human Genetics, University of California, San Francisco, CA*
[4]*Laboratoire d'Anthropobiologie, CNRS-Université Paul Sabatier, Toulouse, France*
[5]*Departamento de Zoología y Antropología Física, Universidad Complutense, Madrid, Spain*
[6]*Départment de Biologie, Faculté des Sciences, Université Chouaïb Doukkali, El Jadida, Morocco*
[7]*Department of Basic Medical Sciences, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates*
[8]*Department of Pathology, Medical School University of Athens, Athens, Greece*

KEY WORDS Berber; genetic admixture; haplotypes; Sahara

ABSTRACT The variation of 18 *Alu* polymorphisms and 3 linked STRs was determined in 1,831 individuals from 15 Mediterranean populations to analyze the relationships between human groups in this geographical region and provide a complementary perspective to information from studies based on uniparental markers. Patterns of population diversity revealed by the two kinds of markers examined were different from one another, likely in relation to their different mutation rates. Therefore, while the *Alu* biallelic variation underlies general heterogeneity throughout the whole Mediterranean region, the combined use of *Alu* and STR points to a considerable genetic differentiation between the two Mediterranean shores, presumably strengthened by a considerable sub-Saharan African genetic contribution in North Africa (around 13% calculated from *Alu* markers). Gene flow analysis confirms the permeability of the Sahara to human passage along with the existence of trans-Mediterranean interchanges. Two specific *Alu*/STR combinations—CD4 110(−) and DM 107(−)—detected in all North African samples, the Iberian Peninsula, Greece, Turkey, and some Mediterranean islands suggest an ancient genetic background of current Mediterranean peoples. Am J Phys Anthropol 000:000–000, 2010. ©2009 Wiley-Liss, Inc.

Following the human population studies of classical markers, the DNA era has been dominated during the last decades by the surveys of uniparental DNA variation (mtDNA and male-specific Y chromosome). In this way, it is interesting to get a complementary perspective from different autosomal independent loci to assure a more accurate approximation to the history of human populations (Pakendorf and Stoneking, 2005; Wilkins, 2006). Consequently, we conducted a study of 18 *Alu* insertion polymorphisms on different chromosomes and 3 microsatellite repeats (STRs) closely linked to some independent *Alu* (CD4, FXIIIB, DM). These analyses have been performed on a wide collection of population samples covering the Mediterranean region. The usefulness of *Alu* polymorphisms and the combination of markers with different mutation rates (*Alu* and STRs) for human population studies has previously been addressed in several examples from other scientific literature (Tishkoff et al., 1996; Watkins et al., 2001; Gaspar et al., 2004; Ramakrishnan and Mountain, 2004; Mateus Pereira et al., 2005; González-Pérez et al., 2007).

The Mediterranean region has played a central role in the genesis and fusion of numerous cultures. This particular geographic area has traditionally been considered an illustrative unit for the study of evolution in modern human populations with geographically and chronologically intertwined elements starting with the Upper Paleolithic cultures (around 30,000 YBP) of North Africa and Europe, which persisted until the Mesolithic (around 8,000 YBP) and extended into the Neolithic age, as testified by numerous prehistoric settlements (Dubief, 1999). This ancient peopling was further fashioned by

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Spanish Ministry of Educación y Ciencia; Grant numbers: CGL2005-03391, CGL2008-03955; Grant sponsor: Generalitat de Catalunya; Grant numbers: 2005SGR00252, 2001FI00177 (to E.G.P.); Grant sponsors: European Science Foundation (E.C. Sixth Framework, EUROCORES Programme, OMLL) via CNRS and Conseil Régional de Midi-Pyrénées, France, and Ministry of Ciencia y Tecnología, Spain; Grant number: ERAS-CT-2003-980409.

*Correspondence to: Pedro Moral, PhD, Secció d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Avda. Diagonal 645, 08028 Barcelona, Spain. E-mail: pmoral@ub.edu

Received 24 March 2009; accepted 9 July 2009

DOI 10.1002/ajpa.21161
Published online in Wiley InterScience (www.interscience.wiley.com).

©2009 WILEY-LISS, INC.

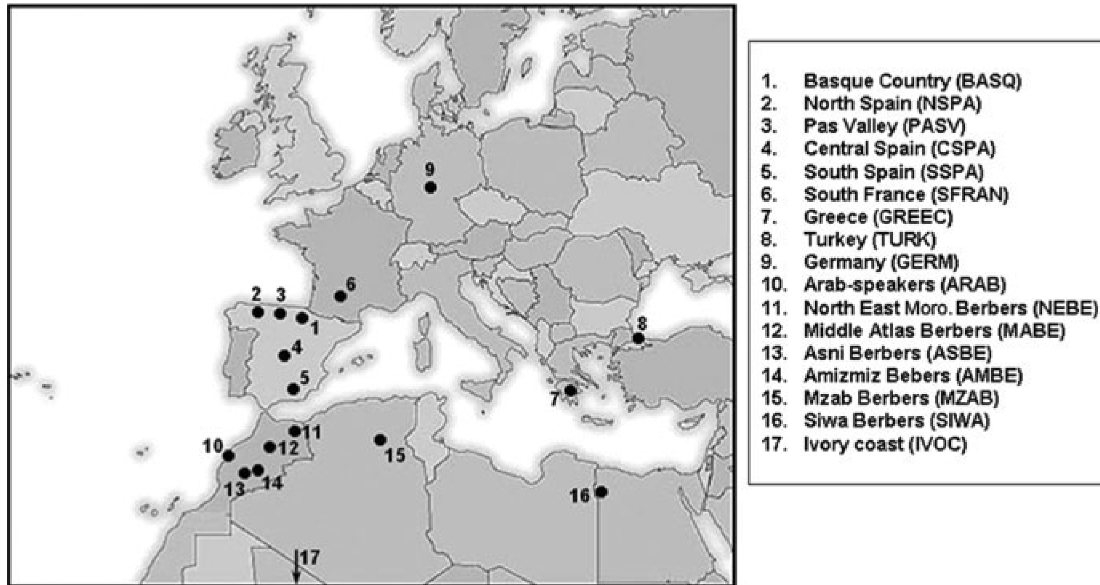

Fig. 1. Geographic location of the populations analyzed in the study.

population movements such as those of the Phoenicians, Greeks, Etruscans, Carthaginians, Romans, and Islamic expansion to North Africa and the Iberian Peninsula, also including the complex migratory waves of the 19th and 20th centuries (Norwich, 2007; Cunliffe, 2008).

As far as the origin of human populations in the Mediterranean is concerned, it is commonly accepted that their roots can be traced back to the Upper Paleolithic with the expansion of human groups from the Near East or Central Asia, or some millennia later with the westward and northward spread of Neolithic populations from the Fertile Crescent. Although there is little doubt regarding the human entrance route to the Mediterranean, controversy appears when different studies try to determine to what extent their current genetic background preserves traces of Paleolithic people and in which degree the almost continuous cultural and political contacts have influenced present genetic affinities. Recent mitochondrial DNA data (Olivieri et al., 2006) suggest a common Levantine source for the Upper Paleolithic cultures that occupied the European (Aurignacian) and North African (Dabban) shores of the Mediterranean. A more recent origin for these populations associated with the demic diffusion of Middle Eastern groups in the Neolithic has been suggested by studies of Y-chromosome (Arredi et al., 2004) and autosomal data (Myles et al., 2005; Tomas et al., 2008). A detailed survey of the E-M78 Y-chromosome haplogroup (Cruciani et al., 2007) indicates the Northeast African origin of this variant and its involvement in trans-Mediterranean migrations from North Africa to Europe during the last 13,000 YBP.

In search of new insights into these questions, this study analyzes a relevant set of Mediterranean populations including eight European samples (from Spain, France, Greece, and Turkey), seven from North Africa (Morocco, Algeria, and Egypt), plus two samples from Central Europe (Germany) and sub-Saharan Africa (Ivory Coast) as external references. Unlike previous surveys of *Alu* variation in the Mediterranean (Comas et al., 2000; González-Pérez et al., 2003), this study deals with a higher number of both *Alu* markers and populations, taking advantage of the combined information derived from *Alu*/STR compound systems.

The main goal of this study is the analysis of relationships among populations in the Mediterranean region based on the data provided by both independent loci and several haplotype systems with fine-scale resolution. This objective was approached by addressing the following: 1) determining the degree of heterogeneity between and within the populations on both shores of the Mediterranean region, 2) comparing population patterns derived from *Alu* polymorphisms with specific signatures from three *Alu*/STR haplotypic systems, all of which can be considered as informative of more recent contacts between populations, 3) detecting past genetic flow through the Mediterranean and the Sahara, and 4) searching for specific *Alu*/STR combinations that could be characteristic of human peopling in the Mediterranean. Our results will be discussed in comparison with those obtained from other genetic markers.

MATERIALS AND METHODS

A total number of 1,831 autochthonous individuals from seven different Mediterranean countries and two samples as external references, Germany (n = 48) in Central Europe and Ivory Coast (n = 121) in sub-Saharan Africa, have been typed. All individuals were healthy blood donors and had all four grandparents born in the same region. Informed consent was obtained from all subjects included in the study, which was approved by the Ethical Committee of the University of Barcelona. Participants were representatives of rural areas of geographically well-defined regions. The geographical location of the recruited samples is shown in Figure 1. The North Mediterranean samples were five Spanish samples [N. Spain (Asturias, n = 114), the Pas Valley (n = 99), the Basque Country (n = 110), Central Spain (Sierra de Gredos, n = 120), and southeast Spain (La Alpujarra, n = 100)] together with South of France (Toulouse, n = 55), Greece (Athica region, n = 96), and Turkey (Anatolia Peninsula, n = 96). The South Mediterranean samples corresponded to several geographical

regions inhabited by well-defined Berber groups from Morocco, Algeria, and Egypt. The Moroccan samples came from High Atlas (Asni, $n = 119$ and Amizmiz Berbers, $n = 86$), Middle Atlas (Berbers from the Khenifra region, $n = 136$), North East Moroccan Berbers (Bouhria area, $n = 111$), plus a sample of Arab-speaking people from the Doukkala region ($n = 98$). Other Berber samples were Mzab from Algeria ($n = 109$) and Siwa Oasis from Egypt ($n = 86$).

Genomic DNA has been extracted from blood by standard phenol–chloroform procedures. Eighteen human-specific *Alu* polymorphic elements (CD4, TPA25, APO, ACE, Yb8NBC120, Yb8NBC125, B65, D1, FXIIIB, A25, PV92, HS2.43, Sb19.3, Sb19.12, HS4.32, HS4.69, DM, Ya5NBC221) together with the Y-chromosome *Alu* polymorphic element (YAP) have been determined. *Alu* genotyping was done by PCR amplification followed by electrophoresis separation using positive and negative controls. Three autosomic STRs have been also determined to reconstruct the haplotype background and evolution of three independent loci: DM, CD4, and FXIIIB. Allele sizes have been verified by sequentiation of different homozygous individuals. PCR amplification conditions and allele nomenclature were according to the protocols described previously (Batzer and Deininger, 1991; Batzer et al., 1996; Brook et al., 1992; Edwards and Gibbs, 1992; Nishimura and Murray, 1992; Arcot et al., 1995; Tishkoff et al., 1996; Comas et al., 2000; Watkins et al., 2001).

Allele frequencies have been calculated by direct gene counting and Hardy–Weinberg equilibrium was checked by an exact test (Guo and Thomson, 1992). Heterozygosity by population and locus has been estimated according to Nei's formula (Nei, 1987). Maximum likelihood haplotype frequencies have been computed using EM algorithm. The geographical structure of the allele frequency variance has been tested by a hierarchical analysis of molecular variation using *F*-Wright statistics from populations clustered according to geographical and/or anthropological criteria. Calculations were performed using the GENEPOP 3.3 (Raymond and Rousset, 1995) and Arlequin (Excoffier et al., 2005) packages.

Population genetic relationships for *Alu* and *Alu*/STR data have been assessed by pairwise F_{ST}-genetic distances (Reynolds et al., 1983) and represented through a multidimensional scaling plot from the distance matrix. Genetic distances $(d\mu)^2$ and R_{ST} values for STR data (Goldstein et al., 1995) have been calculated with the software Microsat2 (written by E. Minch and available at www.hpgl.stanford.edu). A Delaunay network analysis has been performed to identify the main boundaries or regions of sharp genetic discontinuity in the circum-Mediterranean area (Bosch et al., 1997). The isolation degree between populations and, hence, the sense of defining genetic boundaries, has been tested by the ISOLDE program in the GENEPOP package that elucidates if the observed differences could be attributable to an isolation by distance or to abrupt geographical discontinuities preventing human contacts. Levels of genetic admixture have been estimated by means of the maximum likelihood method implemented in the LEADMIX program (Wang, 2003).

Linkage disequilibrium has been calculated using the parameter *d* (Bengtssoon and Thomson, 1981) to approach patterns related with past population movements and expansions. Genetic variation associated with *Alu*+ and *Alu*− chromosomes for each haplotype system has been determined by bootstrapping techniques. The antiquity of some specific Mediterranean haplotypes has been approached according with Slatkin's methodology (Slatkin, 1995). This is based on the Kimura and Ohta's formula (Slatkin and Rannala, 2000) and assumes that the age of an allele can be estimated both from genetic variation among different copies and from its frequency (Kimura and Ohta, 1973). The exact equation, which assumed constant population size and no recombination effects, is $E(t_1) = [(-2p)/(1 - p)]\ln(p)$, where p is the frequency of the neutral variant studied and time is measured in units of $2N$ generations (N is the effective population size, usually assumed as $N = 10{,}000$ during the period before recent growth). For time calculations, it has also been assumed a generation time of 20 years.

The distribution of some characteristic haplotypic frequencies was shaped through a prediction model using an ordinary krigging method implemented in Map-Viewer (GoldenSoftware, CO). Interpolation maps with data collected in this study and from the literature (González-Pérez et al., 2007) were prepared using the same software, and spatial relationships were conceptualized by inverse distance (Relethford, 2008).

TABLE 1. Gene diversities by population and Alu locus

	Population gene diversities		Locus gene diversities	
	Mean heterozygosity	Mean heterozygosity without YAP *Alu*		Mean heterozygosity
BASQ	0.3512	0.3646	*YAP*	0.2313
NSPA	0.3569	0.3641	*DM*	0.4904
PASV	0.3566	0.3493	*HS4.69*	0.4769
CSPA	0.3564	0.3659	*HS4.32*	0.4016
SSPA	0.3584	0.3741	*Alu221*	0.1661
SFRAN	0.3506	0.3614	*Sb19.3*	0.2897
GREEC	0.3633	0.3641	*HS2.43*	0.0526
TURK	0.3526	0.3545	*Sb19.12*	0.3909
GERM	0.3520	0.3661	*B65*	0.4778
ARAB	0.3400	0.3362	*Alu120*	0.4485
NEBE	0.3584	0.3640	*Alu125*	0.2206
MABE	0.3810	0.3871	*PV92*	0.3680
ASBE	0.3641	0.3697	*D1*	0.4419
AMBE	0.3456	0.3609	*FXIIIB*	0.4633
MZAB	0.3383	0.3502	*A25*	0.3053
SIWA	0.3360	0.3357	*CD4*	0.4131
IVOC	0.3567	0.3741	*TPA25*	0.4766
			APOA1	0.1594
			ACE	0.4620

RESULTS

Mediterranean relationships according to *Alu* data

Alu insertion frequencies in the 17 samples analyzed are shown in Supporting Information Table 1. Loci in all samples were in Hardy–Weinberg equilibrium after Bonferroni correction. Locus average heterozygosity across populations ranges from 0.053 (HS2.43) to 0.490 (DM) with around half of loci (53%) showing diversity values close to the maximum expected for biallelic markers (heterozygosities between 0.4 and 0.5). Gene diversity values (Table 1) show statistically significant differences across loci (Kruskal–Wallis test, $P < 0.001$), but not across populations (Kruskal–Wallis test, $P = 0.997$). South and North Mediterranean samples show overlapping frequencies for all markers except for SB19.3 (in Northern Mediterraneans, the frequencies range from 0.862 in Central Spain to 0.972 in the Basque Country,

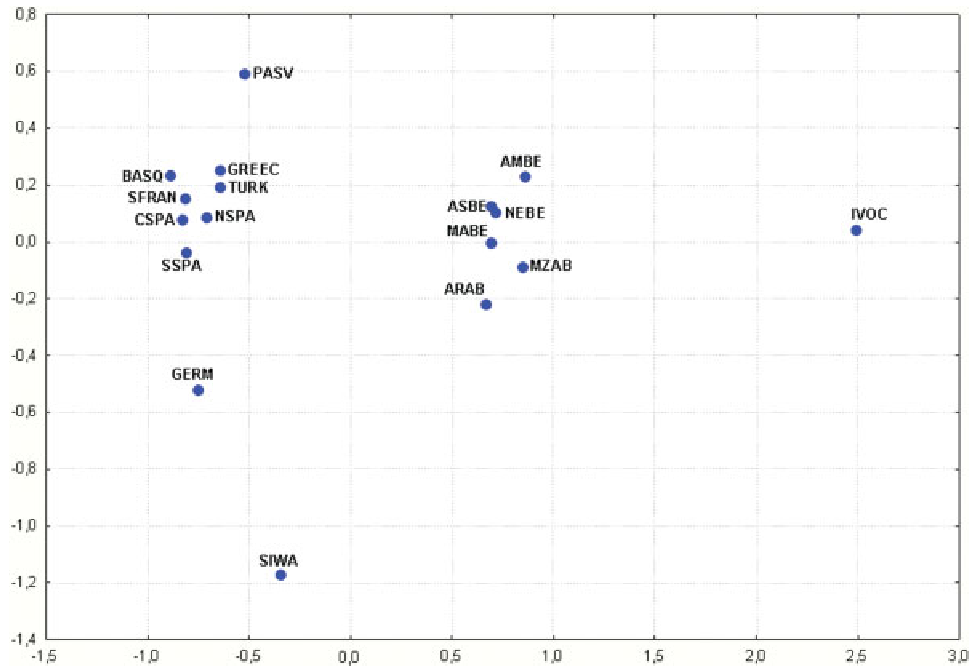

Fig. 2. Multidimensional scaling plot (stress 0.036) applied to the Reynolds' genetic distance matrix based on 18 autosomal *Alu* markers.

whereas in Southern Mediterraneans the frequencies range from 0.614 in Amizmiz to 0.802 in Asni Berbers). Globally, the distribution of autosomal *Alu* loci reveals a remarkable population differentiation (the exact test showed that 130 over a total of 136 pairwise comparisons were statistically significant). The results for Y chromosome YAP insertion confirm the already known dissimilar distribution on both sides of the Mediterranean.

Population relationships estimated by the F_{ST}-related genetic distance index show the greatest distances corresponding to the comparison between sub-Saharans and all others (average distances: sub-Saharan vs. others = 0.166 ± 0.021 and among non-sub-Saharans = 0.040 ± 0.020). In the Mediterranean region, the internal differentiation for Northern Mediterraneans (average distance: 0.022 ± 0.008) is lower than that for Southern Mediterraneans (0.037 ± 0.029). Northern Mediterraneans are almost equidistant from both Central Europe (0.0526 ± 0.015) and Southern Mediterraneans (0.049 ± 0.025), showing distances very close to those between the latter two population groups (0.042 ± 0.013). MDS representation of the genetic distances (see Fig. 2) based on autosomal *Alu* data stresses the main differentiation of sub-Saharans, the clustering of Mediterraneans in two different groups corresponding to northern and southern populations, and the distant position of the Egyptian Siwa and the Spanish Pas Valley samples from their corresponding population clusters. The Siwa oasis sample presents a relatively extreme position, with respect to the other populations. In fact, the first genetic boundary in the Mediterranean separates Siwa Berbers from all remaining groups.

The correlation between geographic and genetic distances under the isolation by distance model yielded highly significant values ($P < 0.002$, after 1,000 bootstrap iterations) suggesting that geographic distance is one of the main factors explaining the genetic variability in this region. This finding is consistent with an analysis of the frequency variance in Mediterraneans. *Alu* markers exhibit a significant degree of variation between populations (6.68%, $P < 0.001$ including the YAP element, and 3.21%, $P < 0.001$ for autosomal *Alu*). However, a hierarchic AMOVA analysis fails to indicate any genetic structure either between East and West (0% among groups vs. 3.25% within-groups variation) or between North and South (2.02% among groups vs. 2.15% within-groups variation).

Population relationships according to STR and *Alu*/STR variation

Allele diversity and some statistical parameters of the allele size distributions of CD4, FXIIIB, and DM microsatellites are reported in Table 2. The three STR distributions were in Hardy–Weinberg equilibrium after Bonferroni correction in all samples. For the CD4 and FXIIIB, Ivory Coast and some North African samples show relatively high diversity than Northern Mediterraneans (see heterozygosity and repeat variance values in Table 2). Analysis of allele size variance indicates a significant interpopulation variation (R_{ST} = 0.0302). Population comparisons also reveal remarkable heterogeneity (112 significant differences out of a total of 136 population comparisons for the overall loci).

Haplotype frequencies for the CD4, FXIIIB, and DM compound systems have been summarized in Supporting Information Tables 2–4. When STR variation has been analyzed separately in *Alu*(+) and *Alu*(−) chromosomes, larger variances are observed in chromosomes carrying the ancestral *Alu* variant: CD4(+), FXIIIB(−), and DM(+). In humans, the ancestral stage of the CD4 and DM loci is the presence of the *Alu* insertion, whereas the absence of the insertion is the ancestral stage for the FXIIIB locus (Brook et al., 1992; Nishimura and Murray, 1992; Tishkoff et al., 1996). *Alu*/STR linkage disequilibrium was present in all systems and samples.

TABLE 2. Variation of CD4, FXIIIB, and DM microsatellites in 17 populations

	CD pentanucleotide					FXIIIB tetranucleotide					DM trinucleotide				
	S. size	*N* Al.	Mean	Coef. var.	Het.	S. size	*N* Al.	Mean	Coef. var.	Het.	S. size	*N* Al.	Mean	Coef. var.	Het.
BASQ	208	5	7.09	0.81	0.646	204	6	8.93	0.17	0.673	205	16	11.07	3.53	0.798
NSPA	196	6	6.82	0.73	0.683	190	6	8.79	0.20	0.711	196	18	10.90	2.81	0.802
PASV	186	4	6.84	0.77	0.629	158	6	8.97	0.13	0.680	186	10	10.52	3.23	0.749
CSPA	184	6	6.99	0.72	0.691	188	5	8.72	0.21	0.721	192	18	10.91	2.95	0.801
SSPA	200	8	7.13	0.78	0.708	182	5	8.84	0.22	0.687	196	15	10.53	2.54	0.777
SFRAN	94	6	7.02	0.80	0.679	94	5	8.97	0.16	0.685	92	16	10.84	3.16	0.790
GREEC	190	6	6.93	0.76	0.686	192	5	8.81	0.14	0.669	192	17	10.74	2.76	0.795
TURK	190	7	7.18	0.82	0.719	190	5	8.83	0.16	0.699	189	20	9.97	2.96	0.748
GERM	92	3	6.42	0.62	0.629	80	6	8.95	0.17	0.705	91	13	10.17	2.10	0.776
ARAB	182	9	7.47	0.86	0.759	148	5	8.21	0.29	0.759	176	14	10.14	2.28	0.756
NEBE	218	8	7.08	0.75	0.722	205	5	8.15	0.25	0.759	218	16	10.27	1.80	0.805
MABE	244	11	7.53	0.91	0.781	242	6	8.16	0.26	0.772	238	15	10.36	2.16	0.789
ASBE	192	11	8.04	0.99	0.786	236	6	8.18	0.28	0.763	237	12	10.02	1.93	0.763
AMBE	170	9	8.09	0.80	0.756	136	6	8.04	0.31	0.781	170	13	9.92	2.21	0.763
MZAB	188	8	7.19	0.84	0.752	186	6	8.09	0.29	0.771	192	11	9.79	1.40	0.756
SIWA	166	9	7.36	0.87	0.764	158	5	7.75	0.31	0.758	172	13	10.41	1.79	0.809
IVOC	190	10	8.28	0.83	0.812	178	6	7.22	0.25	0.695	192	10	10.55	1.22	0.828

S. size: sample size (2N); *N* Al.: number of different alleles detected; Mean: mean of allele size in number of repeats distribution; Coef. var.: coefficient of variation (variance of the allele size distribution divided by the mean); Het.: heterozygosity.

The most obvious pattern of haplotype variation is observed in the CD4 system. The ancestral CD4(+) chromosomes show a decreasing pattern of copy number variation from sub-Saharans to Southern and Northern Mediterraneans. Among these latter populations, the 85(+) and 110(+) haplotypes are the most frequent (Supporting Information Table 2). The derived CD4 *Alu*(−) chromosomes present a lower variation than the ancestral *Alu*(+) chromosomes, which is statistically significant for Northern Mediterraneans ($P < 0.01$) and Southern Mediterraneans ($P < 0.05$), but nonsignificant for the sub-Saharan sample. This reduction trend is considerable in Northern Mediterranean samples (gene diversity: 0.174 for derived chromosomes vs. 0.554 for ancestral ones), moderate in Southern Mediterraneans (0.458 vs. 0.705), and less marked in sub-Saharans (0.721 vs. 0.778).

In contrast, FXIIIB system exhibits general high diversity both in (−) and (+) chromosomes (0.565–0.697 vs. 0.575–0.692, as shown in Supporting Information Table 3). Ancestral *Alu*(−) chromosomes show relatively high frequencies of low copy number alleles (172 and 176) in sub-Saharan population, values higher than 15% of the 172, 184, and 188 variants in Southern Mediterraneans, and a clear predominance of the 188-bp allele (0.319) in Northern Mediterraneans (Supporting Information Table 3). In the derived FXIIIB *Alu*(+) chromosomes, the most frequent allele is the 180-bp variant (>20%) in Mediterraneans, whereas in sub-Saharans the most frequent allele is the 184-bp variant.

In the DM system, ancestral *Alu*(+) chromosomes show a predominance of 77-bp allele in Mediterranean groups (frequencies > 28%). A majority (>24%) of the derived *Alu*(−) chromosomes are carriers of the 98-bp allele in sub-Saharan and Southern Mediterraneans, suggesting that the DM *Alu* deletion probably occurred on a 98-bp variant chromosome (Brook et al., 1992). In Northern Mediterraneans, the most frequent DM(−) variants corresponded to 98- and 101-bp alleles (15.3 and 15.5%, respectively) (see Supporting Information Table 4).

Population differentiation is noticeable in the three haplotype systems (from a total of 136 pairs of comparisons, 102 for CD4, 111 for FXIIIB, and 106 for DM were significant). The genetic variance of *Alu*/STR haplotype frequencies attributable to the differentiation between Northern and Southern Mediterraneans (between groups F_{CT} of 2.2%, $P < 0.001$, and total F_{ST} of 2.9%, $P < 0.001$) indicates significant differences between these two groups. Global population relationships were assessed by genetic distances and depicted in MDS plots: $(d\mu)^2$ for STRs (figure not shown) and F_{ST} for *Alu*/STR haplotypes (see Fig. 3). Both MDS plots revealed a pattern of population relationships very similar to that described for the set of *Alu*. In fact, the three distance matrices showed significant correlations ($P < 0.0001$), with values ranging from 66% (for the comparison between the *Alu* and STR distance matrices) to 87% (for the comparison between STR and *Alu*/STR distance matrices).

Some insights into gene flow and divergence among Mediterraneans

Interestingly, two CD4 and DM particular haplotypes linked to *Alu*-derived alleles—CD4 110(−) and DM 107(−)—were relatively frequent in Berber samples, present at lower frequencies in some Northern Mediterraneans, but absent in sub-Saharans and Central Europeans (see frequency distributions in Supporting Information Tables 2–4). The CD4 110(−) haplotype showed polymorphic frequencies only in Northern Mediterraneans from Northern Spain (2.7%) and Southern France (2.2%), with values in North Africa ranging between 7% (High Atlas Amizmiz Berbers) and 1.5%. This haplotype was not detected in Algerian Mozabite Berbers (see Fig. 4A). The DM 107(−) haplotype presented polymorphic frequencies in Central Spain (2.2%), Greece (1.0%), and Turkey (1.0%). In North Africa, this combination was detected in all samples with the lowest values in Moroccan Arab-speakers and Middle Atlas Berbers (1.5%), values around 2.2% in Mozabite and Siwa Berbers, and the highest frequencies (5%) in North East Moroccan Berbers and High Atlas Asni and Amizmiz Berbers (see Fig. 4B). Age estimates of these two haplotypes yielded values of 36,000 (95% CI: 3,200–186,000) YBP for the CD4 110(−) and 48,000 (95% CI: 4,550–265,000) YBP for the DM 107(−) combination.

TABLE 3. Maximum likelihood estimations of sub-Saharan African genetic contribution (P1) to the North and South Mediterranean gene pools

	18 Autosomal *Alus*			3 *Alu*/STR compound systems		
	Mode	+95% CI	−95% CI	Mode	+95% CI	−95% CI
Southern Mediterraneans[a]						
Moroccan Arabs	0.0645	0.2959	<0.001	0.3437	0.5140	0.1820
North East Atlas Berbers	0.0682	0.2424	<0.001	0.1678	0.2978	0.0606
Middle Atlas Berbers	0.1249	0.3041	<0.001	0.3245	0.4810	0.1974
High Atlas Asni Berbers	0.0641	0.2578	<0.001	0.3338	0.5258	0.1573
High Atlas Amizmiz Berbers	0.1684	0.4074	<0.001	0.3532	0.5500	0.2006
Mozabite Berbers	0.0996	0.5419	<0.001	0.3769	0.5501	0.2357
Siwa Berbers	0.0005	0.3134	<0.001	0.5109	0.6772	0.3218
South Mediterraneans as a whole	0.1297	0.3267	<0.001	0.3949	0.5423	0.2954
Northern Mediterraneans[b]						
Basque Country	0.0003	0.1918	<0.001	0.0651	0.1785	<0.001
North Spain	0.0002	0.1341	<0.001	0.1108	0.2050	0.0406
Pas Valley	0.0001	0.1320	<0.001	0.0776	0.2043	<0.001
Central Spain	0.0003	0.1459	<0.001	0.1500	0.2692	0.0494
South Spain	0.0004	0.3160	<0.001	0.1253	0.2392	0.0526
South France	0.0003	0.1443	<0.001	0.1253	0.2353	0.0433
Greece	0.0001	0.1305	<0.001	0.0835	0.1822	0.0168
Turkey	0.0002	0.1353	<0.001	0.0938	0.2033	0.0185
North Mediterraneans as a whole	0.0001	0.0891	<0.001	0.0501	0.0782	0.0250

[a] Parental 1: sub-Saharan Africans, Parental 2: Northern Mediterraneans.
[b] Parental 1: sub-Saharan Africans, Parental 2: Central Europeans.

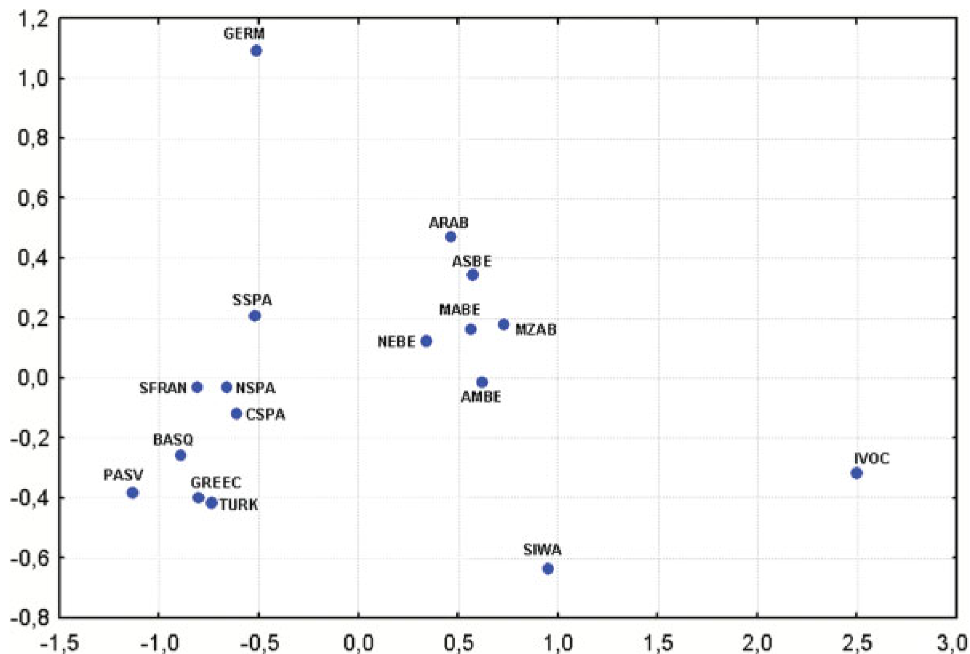

Fig. 3. Multidimensional scaling plot (stress 0.049) applied to the Reynolds' genetic distance matrix based on three *Alu*/STR compound systems.

Gene flow in Mediterraneans (North and South shores) was analyzed with LEADMIX simulations checking for different parental groups. The only consistent results had been summarized in Table 3 (admixture based on STRs are almost identical than those obtained from *Alu*/STR haplotypes and, hence, they are not included in Table 3). In the case of Southern Mediterraneans, the overall sub-Saharan contribution ranges from 13% for the *Alu* markers to 46 and 40% for the STRs and *Alu*/STR data, respectively. The sub-Saharan contribution in Northern Mediterraneans was imperceptible for the *Alu* data set, but reaches values around 6% for the STRs set and *Alu*/STR haplotypes.

As for individual populations, the sub-Saharan gene flow in North Africa based on the *Alu* data collection ranges between 6 and 17% (Table 3), except the Siwa Berbers where that influence was negligible. Admixture values based on *Alu*/STR combinations indicate that sub-Saharan flow in North Africa ranged from 16% (North East Moroccan Berbers) to 35% (remaining samples) with the exception of Siwa Berbers who showed the highest admixture value (51%).

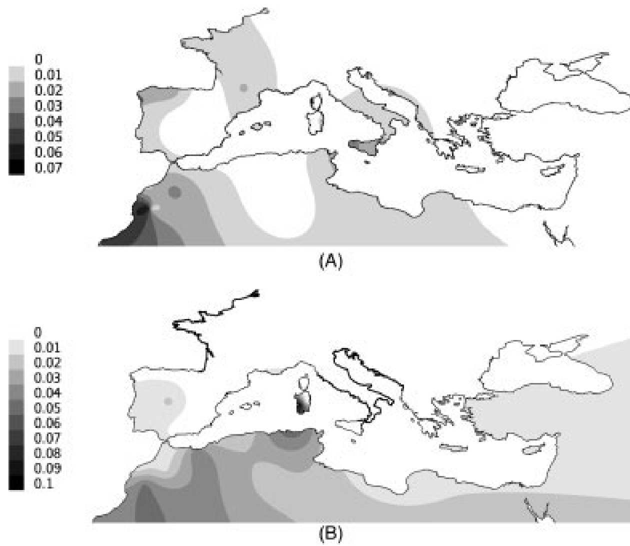

Fig. 4. Frequency distribution maps of haplotypes CD4 110(−) (**A**) and DM 107(−) (**B**).

Concerning Northern Mediterraneans, the gene flow from sub-Saharan Africa was inappreciable for *Alu* markers and swung from 6 to 15% for the *Alu*/STR haplotypes data calculations. When gene flow in Northern Mediterraneans was tested, taking Central Europe and Southern Mediterraneans as parental populations, the results were statistically inconsistent, indicating the limited power of our markers to discriminate gene flow within Caucasoid populations. Nonetheless, the distributions of frequencies for the Mediterranean haplotypes CD4 110(−) and DM 107(−) (Fig. 4A,B) are suggestive of gene flow processes across this geographical region.

DISCUSSION

This autosomal genetic study carried out in 15 anthropologically well-defined Mediterranean populations was designed to analyze population relationships and give complementary insight on that derived from previous genetic studies (Quintana-Murci et al., 2003; Esteban et al., 2006; Gerard et al., 2006; Olivieri et al., 2006; González-Pérez et al., 2007; Tomas et al., 2008).

The high correlation found between genetic distances from the three sets of markers (*Alu*, STR, and *Alu*/STR haplotypes) demonstrates that the relationship pattern derived from the autosomal polymorphisms analyzed is consistent. This pattern identifies Mediterranean populations as genetically separate from both sub-Saharans and Central Europeans and allows the identification of a certain genetic structure between the two shores of the Mediterranean region. This genetic picture of populations may be related to geographic factors as indicated by the high correlation ($P < 0.002$) between geographic and genetic distances (based on *Alu* markers) found under the isolation by distance model. The genetic distinctiveness of Mediterranean populations, as well as the distinction between Northern and Southern Mediterraneans, coincides with results in previous studies (see for instance, Simoni et al., 1999; Comas et al., 2000; Bosch et al., 2001). In this general view, it is worth noting the particular position of two populations (the Spanish Pas Valley and the Egyptian Siwa Berbers) (see Fig. 2). These two populations have previously been described as genetic outliers (Esteban et al., 2006; Moral et al., 2006; Coudray et al., 2009) due to the orography of the Pas Valley and the desert surrounding the Siwa Oasis. This isolation could explain their differentiation by the action of the genetic drift associated with episodes with low effective population size, which in the case of Siwa Oasis, could have enhanced the effect of sub-Saharan flow (51% from *Alu*/STR data) through the Nile River (Fakhry, 1973).

The variance of the autosomal *Alu* loci frequencies indicates significant genetic heterogeneity in the Mediterranean region ($F_{ST} = 3.21\%$, $P < 0.001$) in contrast to other *Alu*-based studies that emphasize the homogeneity of this area, in particular the northern side (Comas et al., 2000; Tomas et al., 2008). These discrepancies are likely associated with the different number of markers used. The *Alu* polymorphisms of this study also evidence significant genetic diversity in the northern and southern shores of the Mediterranean (North: $F_{ST} = 1.6\%$, $P < 0.001$; South: $F_{ST} = 2.8\%$, $P < 0.001$). The heterogeneity is maintained even after the assumed outlying Pas Valley and Siwa samples are removed from calculations (F_{ST} of 1.4%, $P < 0.001$ in each shore). Another remarkable finding is the comparatively higher heterogeneity in the southern part of the Mediterranean. This can be explained by longer periods of relative isolation and low population sizes that had been reported elsewhere for North African populations in comparison with the Northern Mediterraneans (Fadhlaoui-Zid et al., 2004).

When we test the consistency of the Mediterranean as a genetic barrier, the apportionment of *Alu* genetic variance failed to recognize any geographic structure for the population genetic diversity: variation between North and South (F_{CT} of 2.02%, $P < 0.001$) is slightly lower than that observed within groups (F_{SC} of 2.15% $P < 0.001$). In contrast, the hierarchical AMOVA analysis of the variation of STR and *Alu*/STR haplotypes is consistent with a geographic structuring of the population gene diversity. In fact, the majority of variance corresponds to differences between North and South (within groups F_{SC} of 0.7%, $P < 0.01$, and between groups F_{CT} of 2.2%, $P < 0.001$) as compared with the variation within these geographic groups.

The disparity between the results from *Alu* loci and *Alu*/STR haplotypes, apart from the potential effect of the different number of independent markers examined (18 vs. 3), could be related to different mutation rates and therefore the power to detect ancient or more recent demographic events. Similar disparities between these two kinds of markers were found in the admixture analysis (Table 3).

As far as gene flow analysis is concerned, the variability in the markers considered and the statistical methods available allow the quantification of sub-Saharan gene flow into the Mediterranean region as indicated in Table 3. The estimates of sub-Saharan gene flow in Southern Mediterraneans oscillated between 12.9% (*Alu* loci) and 39.5% (*Alu*/STR haplotypes), a wide range probably related with the different mutational nature of the markers analyzed and with the effect of repeated homoplasic mutation in STRs. The presence of sub-Saharan African traces in the gene pool of North Africans supports the idea of the permeability of the Sahara desert to human migrations as reported in other studies for different kinds of markers (see for example, Plaza

et al., 2003; Arredi et al., 2004; Myles et al., 2005; Coudray et al., 2006). Interestingly, data from mtDNA and Y-chromosome estimates of sub-Saharan gene flow in North Africa are similar to that obtained from our *Alu* loci set, a value also concordant with that corresponding to Mozabites in the recent survey of Li et al. (2008) based on more than 500,000 SNPs. The interpretation of the disparity in gene flow estimates according to the kind of marker is difficult, but it might be presumably be related to the different mutation rates of *Alu* and STRs. In any case, the aforementioned consistency with other studies and markers and the reduced number of loci analyzed lead to think that *Alu*/STR estimate might be artefactual.

Regarding gene flow within the Mediterranean region and unlike the unilinearly transmitted regions of the genome, the autosomal markers analyzed do not have enough power to discern gene flow estimates from south to north and between western and eastern parts of the region. Nevertheless, our findings include indirect indicators of the gene flow that occurred throughout the Mediterranean region. In Northern Mediterraneans, traces from sub-Saharan African genes were detected through admixture analysis (Table 3) suggesting continuous contacts among both Mediterranean shores as it has been also described in other reports (González-Pérez et al., 2003; Plaza et al., 2003). The fact that these traces have been detected in the entirety of the northern shore, from Spain (10%) to Turkey (9.4%), reinforces the hypothesis that gene flow in this region is probably linked to the first ancient trans-Mediterranean navigations and that has been maintained and homogenized through enduring slave trade that prevailed until the end of the 17th century (Olesa-Muñido, 1968), rather than merely reflecting the Islamic expansions of the 8th to 15th centuries.

Another indication of gene flow across the Mediterranean comes from the distribution of specific *Alu*/STR combinations such as CD4 110(−) and DM 107(−) haplotypes. So far, these haplotype combinations almost exclusively detected in Mediterranean human populations testify to the existence of specific Mediterranean trends according to previous records (Esteban et al., 2004). Although the similar age estimates of 36,000 (95% CI: 3,200–186,000) YBP for the CD4 110(−) and 48,000 (95% CI: 4,550–265,000) YBP for the DM 107(−) suggest an ancient and coincident origin in the region, the wide range of confidence intervals and the assumptions used by the method does not allow clear statements.

The highest frequencies of CD4 110(−) and DM 107(−) have been found in the High Atlas region (7 and 5.5%, respectively) of Morocco, reaching polymorphic frequencies in all the North African samples [barring the Mozabites for the CD4 110(−) combination]. They have also been found in the Iberian Peninsula, scattered along the northern Mediterranean shore to Greece and Turkey, and on the main islands of the western Mediterranean (Majorca, Corsica, Sardinia, and Sicily; González-Pérez et al., 2007). The CD4 110(−) haplotype has also been reported in West Saharans and Mauritanians (Flores et al., 2000) and on five of the seven Canary Islands (Flores et al., 2001), as well as in Adygei from the Northern Caucasus (Tishkoff et al., 1996).

Assuming from their frequency distribution that the place of origin of these particular haplotypes is located in the westernmost extreme of North Africa (Fig. 4A,B), their current ample distribution along both shores of the Mediterranean most likely reflects the effect of gene flow across the region since ancient times, even though specific ages cannot be accurately estimated with our data. Similarly, specific Mediterranean haplogroups or clades (U6 and M1b in the mtDNA; EM78 and EM81 in the Y-chromosome) have also been described for these populations and dated in Paleolithic times.

In summary, the population information from autosomal data concurs with studies based on uniparental markers (see for example, Plaza et al., 2003; Achilli et al., 2004; Olivieri et al., 2006; Cruciani et al., 2007). The results of this study are consistent with the following statements: i) human populations in the Mediterranean present distinctive genetic features from sub-Saharan and Central European groups; ii) the degree of heterogeneity in this region is statistically significant, being relatively higher in Southern Mediterraneans, likely in association with their demographic history; iii) Mediterranean gene diversity is compatible with ascertained (but only slight) differentiation North vs. South, to which the different sub-Saharan gene flow has likely played an important role; iv) the distribution of some specific Mediterranean *Alu*/STR haplotypes is indicative of a history of gene flow interactions across this geographic region. In any case, this study demonstrates the utility of combined use of autosomal *Alu* and STR polymorphisms to detect fine population relationships. Differences in gene flow estimates obtained from the different groups of markers used strengthened the necessity of taken into account the particular mutational nature to explain properly human population events. The general agreement shown by uniparental and autosomal data confirms the capacity of human population genetics to reconstruct, through different tools, the same population history-in this case, that of the Mediterraneans.

ACKNOWLEDGMENTS

We thank all the anonymous people for their participation in the study and Dr. María del Carmen Rodríguez and Dr. Mauricio De Grado for their contribution to the sampling. We would also like to give special thanks to Christopher Gignoux (University of California, San Francisco) for his help with MapViewer software and to Anthony D. Loera for the careful English revision of the manuscript. The authors highly appreciate the helpful comments and suggestions made by the editors and two anonymous reviewers.

LITERATURE CITED

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75:910–918.

Arcot SS, Shaikh TH, Kim J, Bennett L, Alegria-Hartman M, Nelson DO, Deininger PL, Batzer MA. 1995. Sequence diversity and chromosomal distribution of "young" Alu repeats. Gene 163:273–278.

Arredi B, Poloni ES, Parachini S, Zerjal T, Fathallah DM, Makfrelouf M, Pascali VL, Novelletto A, Tyler-Smith C. 2004. A predominantly neolithic origin for Y-chromosome variation in North Africa. Am J Hum Genet 75:338–345.

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.

Batzer MA, Deininger PL. 1991. A human-specific subfamily of Alu sequences. Genomics 9:481–487.

Bengtssoon BO, Thomson G. 1981. Measuring the strength of associations between HLA antigens and diseases. Tissue Antigens 18:356–363.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am J Hum Genet 68:1019–1029.

Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, Bertranpetit J. 1997. Population history of north Africa: evidence from classical genetic markers. Hum Biol 69:295–311.

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T. 1992. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68:799–808.

Comas D, Calafell F, Benchemsi N, Helel A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Strait. Hum Genet 107:312–319.

Coudray C, Guitard E, Kandil M, Harich N, Melhaoui M, Baali A, Sevin A, Moral P, Dugoujon JM. 2006. Study of GM immunoglobulin allotypic system in Berbers and Arabs from Morocco. Am J Hum Biol 18:23–34.

Coudray C, Olivieri A, Achilli A, Maria Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossman M, Torroni A, Dugoujon JM. 2009. The complex and diversified mitochondrial gene pool of Berber populations. Ann Hum Genet 73:196–214.

Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellito D, Colomb EB, Dugoujon JM, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradisnik B, Assum G, Brdicka R, Kozlov AI, Efremov GD, Coppa A, Novelletto A, Scozzari R. 2007. Tracing past human male movements in Northern/Eastern Africa and West Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. Mol Biol Evol 24:1300–1311.

Cunliffe B. 2008. Europe between the oceans: 9000 BC–AD 1000. New Haven: Yale University Press.

Dubief J. 1999. L'Ajjer, Sahara central. Paris: Karthala Editions.

Edwards MC, Gibbs RA. 1992. A human dimorphism resulting from loss of an Alu. Genomics 14:590–597.

Esteban E, González-Pérez E, Harich N, López-Alomar A, Via M, Luna F, Moral P. 2004. Genetic relationships among Berbers and South Spaniards based on CD4 microsatellite/Alu haplotypes. Ann Hum Biol 31:202–212.

Esteban E, Rodon N, Via M, González-Pérez E, Santamaria J, Dugoujon JM, El Chennawi F, Melhaoui M, Cherkaoui M, Vona G, Harich N, Moral P. 2006. Androgen receptor CAG and GGC polymorphisms in Mediterraneans: repeat dynamics and population relationships. J Hum Genet 51:129–136.

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

Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar El gaaied A. 2004. Mitochondrial DNA heterogeneity in Tunisian Berbers. Ann Hum Genet 68:222–233.

Fakhry A. 1973. Siwa oasis. Cairo: The American University in Cairo Press.

Flores C, Maca-Meyer N, Gonzalez AM, Cabrera VM. 2000. Northwest African distribution of the CD4/Alu microsatellite haplotypes. Ann Hum Genet 64:321–327.

Flores C, Maca-Meyer N, Perez JA, Cabrera VM. 2001. The peopling of the Canary Islands: a CD4/Alu Microsatellite haplotype perspective. Hum Immunol 62:949–953.

Gaspar P, Seixas S, Rocha J. 2004. Genetic variation in a compound short tandem repeat/Alu haplotype system at the SB19.3 locus: properties and Interpretation. Hum Biol 76:277–287.

Gerard N, Berriche S, Aouizerate A, Dieterlen F, Lucotte G. 2006. North African Berber and Arab influences in the Western Mediterranean revealed by Y-chromosome DNA haplotypes. Hum Biol 78:307–316.

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW. 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471.

González-Pérez E, Moral P, Via M, Vona G, Varesi L, Santamaria J, Gayà-Vidal M, Esteban E. 2007. The ins and outs of population relationships in West Mediterranean Islands: data from autosomal Alu polymorphisms and Alu/STR compound systems. J Hum Genet 52:999–1010.

González-Pérez E, Via M, Esteban E, López-Alomar A, Mazieres S, Harich N, Kandil M, Dugoujon JM, Moral P. 2003. Alu insertions in the Iberian Peninsula and NW Africa: genetic boundaries or melting pot? Coll Antropol 27:491–500.

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

Kimura M, Ohta T. 1973. The age of a neutral mutant persisting in a finite population. Genetics 75:199–212.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104.

Mateus Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM, Herrera RJ. 2005. Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. Am J Phys Anthropol 128:171–184.

Moral P, Esteban E, DeGrado M. 2006. Genetics of a human cultural and geographic isolate: the Pas Valley population from the North of the Iberian Peninsula. Human Genetic Isolates: 29–37. Carla Caló and Giuseppe Vona Ed.

Myles S, Bouzekri N, Haverfield E, Cherkaoui M, Dugoujon JM, Ward R. 2005. Genetic evidence in support of a shared Eurasian-North African dairying origin. Hum Genet 117:34–42.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nishimura DY, Murray JC. 1992. A tetranucleotide repeat for the F13B locus. Nucleic Acid Res 20:1167.

Norwich JJ. 2007. The Middle Sea: a history of the Mediterranean. London: Secker and Warburg.

Olesa Muñido F. 1968. La organización naval de los estados mediterráneos, y en especial de España durante los siglos XVI y XVII. Madrid.

Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt HJ, Torroni A. 2006. The mtDNA legacy of the Levantine Early Upper Palaeolithic in Africa. Science 314:1767–1770.

Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. Annu Rev Genomics Hum Genet 6:497–429.

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D. 2003. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean. Ann Hum Genet 67:312–328.

Quintana-Murci L, Veitia R, Fellous M, Semino O, Poloni ES. 2003. Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. Am J Phys Anthropol 121:157–171.

Ramakrishnan U, Mountain JL. 2004. Precision and accuracy of divergence time estimates from STR and SNPSTR variation. Mol Biol Evol 21:1960–1971.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249.

Relethford JH. 2008. Geostatistics and spatial analysis in biological anthropology. Am J Phys Anthropol 136:1–10.

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

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.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Slatkin M, Rannala B. 2000. Estimating allele age. Annu Rev Genomics Hum Genet 1:225–49.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380–1387.

Tomas C, Sanchez JJ, Barbaro A, Brandt-Casadevall C, Hernandez A, Ben Dhiab M, Ramon M, Morling N. 2008. X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans). BMC Evol Biol 8:75.

Wang J. 2003. Maximum-likelihood estimation of admixture proportions from genetic data. Genetics 164:747–765.

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

Wilkins JF. 2006. Unraveling male and female histories from human genetic data. Curr Opin Genet Dev 16:611–617.

Table S1a. *Alu* insertion frequencies and heterozygosity values in 17 populations.

	BASQ	NSPA	PASV	CSPA	SSPA	SFRAN	GREEC	TURK	GERM	ARAB	NEBE	MABE	ASBE	AMBE	MZAB	SIWA	IVOC
YAP																	
N	52	77	47	20	51	36	54	36	20	43	53	56	52	28	60	51	45
Insertion	0.058	0.130	0.404	0.100	0.039	0.083	0.222	0.194	0.050	0.721	0.849	0.839	0.846	0.964	0.933	0.216	0.978
Het	0.110	0.228	0.487	0.185	0.076	0.155	0.349	0.318	0.097	0.407	0.259	0.272	0.263	0.07	0.126	0.342	0.044
DM																	
N	104	89	77	89	97	46	90	96	39	49	94	120	119	74	87	70	91
Insertion	0.548	0.466	0.578	0.584	0.438	0.587	0.544	0.573	0.308	0.500	0.479	0.383	0.496	0.487	0.368	0.357	0.539
Het	0.498	0.501	0.491	0.489	0.495	0.490	0.499	0.492	0.432	0.505	0.502	0.475	0.502	0.503	0.468	0.463	0.500
HS4.69																	
N	91	93	86	92	95	54	91	96	34	84	90	94	96	81	80	64	89
Insertion	0.456	0.436	0.343	0.386	0.432	0.426	0.407	0.406	0.368	0.381	0.367	0.394	0.474	0.389	0.288	0.359	0.365
Het	0.499	0.494	0.453	0.477	0.493	0.494	0.485	0.485	0.472	0.475	0.467	0.480	0.501	0.478	0.412	0.464	0.466
HS4.32																	
N	89	112	67	89	95	48	94	95	27	92	90	97	96	81	83	72	91
Insertion	0.601	0.692	0.649	0.494	0.679	0.708	0.766	0.811	0.519	0.761	0.767	0.644	0.787	0.71	0.645	0.632	0.319
Het	0.482	0.428	0.459	0.503	0.438	0.418	0.361	0.309	0.509	0.366	0.360	0.461	0.338	0.415	0.461	0.468	0.437
Alu221																	
N	91	95	91	88	40	48	96	95	30	90	90	106	96	81	101	81	90
Insertion	0.962	0.979	0.978	0.972	0.725	0.969	0.974	0.911	0.917	0.911	0.911	0.887	0.891	0.889	0.876	0.753	0.767
Het	0.074	0.041	0.043	0.056	0.404	0.061	0.051	0.164	0.155	0.163	0.163	0.202	0.196	0.199	0.218	0.374	0.360
Sb19.3																	
N	107	100	72	87	68	48	95	96	40	93	90	88	96	79	90	81	85
Insertion	0.972	0.880	0.924	0.862	0.882	0.938	0.879	0.906	0.675	0.704	0.778	0.688	0.802	0.614	0.683	0.654	0.177
Het	0.055	0.212	0.142	0.239	0.209	0.118	0.214	0.171	0.444	0.419	0.348	0.432	0.319	0.477	0.435	0.455	0.292
HS2.43																	
N	107	99	99	120	96	48	95	96	42	86	103	123	119	86	107	82	114
Insertion	0.108	0.081	0.035	0.100	0.099	0.083	0.068	0.037	0.012	0.029	0.019	0.041	0.042	0.041	0.023	0.006	0.000
Het	0.193	0.149	0.069	0.181	0.179	0.154	0.128	0.071	0.024	0.057	0.038	0.078	0.081	0.079	0.046	0.012	0.000
Sb19.12																	
N	107	99	99	120	96	48	96	96	48	78	90	121	96	86	107	76	113
Insertion	0.402	0.364	0.288	0.400	0.224	0.365	0.354	0.260	0.302	0.122	0.206	0.260	0.245	0.163	0.136	0.283	0.394
Het	0.483	0.465	0.412	0.482	0.349	0.468	0.460	0.387	0.426	0.215	0.328	0.387	0.372	0.274	0.235	0.408	0.480
B65																	
N	104	99	95	120	100	48	93	96	41	94	103	122	118	83	99	63	113
Insertion	0.606	0.500	0.411	0.579	0.540	0.563	0.608	0.578	0.512	0.505	0.461	0.545	0.619	0.621	0.525	0.151	0.690
Het	0.480	0.503	0.487	0.490	0.499	0.497	0.480	0.490	0.506	0.503	0.499	0.498	0.474	0.474	0.501	0.258	0.430

Table S1b. *Alu* insertion frequencies and heterozygosity values in 17 populations.

	BASQ	NSPA	PASV	CSPA	SSPA	SFRAN	GREEC	TURK	GERM	ARAB	NEBE	MABE	ASBE	AMBE	MZAB	SIWA	IVOC
Alu120																	
N	109	99	99	114	96	48	96	96	48	83	96	122	96	86	109	86	94
Insertion	0.473	0.465	0.369	0.272	0.469	0.417	0.406	0.453	0.385	0.241	0.505	0.459	0.516	0.57	0.381	0.023	0.739
Het	0.501	0.500	0.468	0.398	0.501	0.491	0.485	0.498	0.479	0.368	0.503	0.499	0.502	0.493	0.474	0.046	0.388
Alu125																	
N	110	100	99	110	96	48	96	95	48	83	96	122	96	85	109	84	96
Insertion	0.114	0.130	0.359	0.127	0.130	0.198	0.219	0.174	0.073	0.048	0.099	0.143	0.104	0.147	0.078	0.066	0.037
Het	0.202	0.227	0.462	0.223	0.228	0.321	0.344	0.289	0.137	0.092	0.179	0.247	0.188	0.252	0.145	0.123	0.071
PV92																	
N	108	99	66	120	100	48	94	96	39	76	103	122	118	82	99	82	119
Insertion	0.185	0.141	0.076	0.213	0.185	0.240	0.298	0.313	0.256	0.428	0.34	0.369	0.424	0.146	0.369	0.256	0.143
Het	0.303	0.244	0.141	0.336	0.303	0.368	0.421	0.432	0.386	0.493	0.451	0.468	0.5	0.251	0.468	0.383	0.246
D1																	
N	106	99	66	120	100	48	96	96	42	53	103	121	119	84	107	83	120
Insertion	0.481	0.394	0.439	0.383	0.325	0.396	0.396	0.375	0.381	0.226	0.286	0.298	0.336	0.292	0.210	0.217	0.388
Het	0.502	0.480	0.496	0.475	0.441	0.483	0.481	0.471	0.477	0.354	0.411	0.420	0.448	0.416	0.334	0.342	0.477
FXIII																	
N	107	98	96	119	100	48	96	95	34	61	111	122	119	81	102	83	117
Insertion	0.514	0.429	0.526	0.395	0.380	0.365	0.568	0.584	0.162	0.230	0.396	0.398	0.340	0.315	0.348	0.295	0.333
Het	0.502	0.492	0.501	0.480	0.474	0.468	0.493	0.488	0.275	0.357	0.481	0.481	0.451	0.434	0.456	0.419	0.446
A25																	
N	91	100	96	120	100	48	96	95	47	65	109	122	118	81	106	85	118
Insertion	0.220	0.170	0.156	0.175	0.085	0.156	0.115	0.100	0.181	0.062	0.133	0.123	0.119	0.111	0.142	0.165	0.352
Het	0.345	0.284	0.265	0.290	0.156	0.266	0.204	0.181	0.300	0.116	0.232	0.217	0.210	0.199	0.244	0.277	0.458
CD4																	
N	107	110	88	94	99	54	96	95	34	98	100	116	103	82	92	50	116
Insertion	0.804	0.636	0.852	0.702	0.702	0.741	0.719	0.737	0.603	0.658	0.620	0.612	0.660	0.659	0.685	0.740	0.767
Het	0.317	0.465	0.253	0.421	0.421	0.388	0.406	0.390	0.486	0.452	0.474	0.477	0.451	0.453	0.434	0.389	0.359
TPA25																	
N	102	114	95	95	98	47	96	96	43	98	102	136	118	83	107	82	121
Insertion	0.588	0.526	0.563	0.574	0.526	0.457	0.578	0.474	0.372	0.490	0.662	0.485	0.644	0.615	0.463	0.317	0.227
Het	0.487	0.501	0.495	0.492	0.501	0.502	0.490	0.501	0.473	0.502	0.450	0.501	0.460	0.477	0.500	0.436	0.353
APOA1																	
N	108	113	87	86	97	54	96	96	45	83	103	116	119	85	102	50	121
Insertion	0.926	0.960	0.920	0.971	0.923	0.982	0.953	0.953	0.944	0.88	0.864	0.892	0.857	0.894	0.941	0.840	0.376
Het	0.138	0.077	0.149	0.057	0.143	0.037	0.090	0.090	0.106	0.213	0.236	0.193	0.246	0.191	0.111	0.272	0.471
ACE																	
N	100	114	93	92	100	55	96	95	47	92	90	135	94	86	102	86	121
Insertion	0.485	0.421	0.479	0.467	0.46	0.391	0.359	0.379	0.479	0.277	0.311	0.344	0.293	0.314	0.235	0.343	0.484
Het	0.502	0.490	0.502	0.501	0.499	0.481	0.463	0.473	0.505	0.403	0.431	0.453	0.416	0.433	0.362	0.453	0.502

Table S2. CD4 *Alu*/STR haplotype frequencies and diversity statistics in 17 populations. STR alleles are expressed in base pairs. Pooled frequencies are also given for the European (9 samples) and North African (7 samples) groups.

STR alleles	**75**	**80**	**85**	**90**	**95**	**100**	**105**	**110**	**115**	**120**	**125**	**130**	**135**
					Haplotypes with the Alu *insertion (*Alu+*)*								
BASQ	0.000	0.000	0.426	0.010	0.000	0.000	0.000	0.343	0.024	0.010	0.000	0.000	0.000
NSPA	0.000	0.000	0.387	0.000	0.000	0.000	0.000	0.212	0.023	0.016	0.000	0.000	0.000
PASV	0.000	0.000	0.501	0.000	0.000	0.000	0.000	0.308	0.023	0.000	0.000	0.000	0.000
CSPA	0.000	0.000	0.372	0.011	0.000	0.000	0.011	0.294	0.022	0.000	0.000	0.000	0.000
SSPA	0.000	0.000	0.343	0.020	0.005	0.000	0.015	0.268	0.030	0.005	0.000	0.015	0.000
SFRAN	0.000	0.000	0.402	0.000	0.000	0.000	0.000	0.283	0.022	0.011	0.011	0.000	0.000
GREEC	0.000	0.000	0.389	0.021	0.000	0.000	0.000	0.262	0.032	0.005	0.000	0.005	0.000
TURK	0.000	0.000	0.376	0.011	0.000	0.000	0.011	0.250	0.081	0.000	0.005	0.000	0.000
GERM	0.000	0.000	0.300	0.033	0.000	0.000	0.000	0.288	0.000	0.000	0.000	0.000	0.000
EUROPEANS	0.000	0.000	0.391	0.011	0.001	0.000	0.005	0.277	0.032	0.006	0.001	0.003	0.000
ARAB	0.000	0.000	0.225	0.044	0.008	0.011	0.011	0.212	0.111	0.012	0.000	0.011	0.000
NEBE	0.000	0.000	0.311	0.000	0.000	0.016	0.000	0.243	0.030	0.010	0.000	0.010	0.000
MABE	0.005	0.000	0.022	0.016	0.000	0.038	0.005	0.175	0.114	0.006	0.000	0.019	0.000
ASBE	0.000	0.000	0.183	0.034	0.000	0.015	0.010	0.220	0.095	0.032	0.010	0.053	0.000
AMBE	0.000	0.000	0.223	0.014	0.000	0.019	0.012	0.263	0.085	0.025	0.012	0.007	0.000
MZAB	0.000	0.000	0.312	0.023	0.000	0.031	0.000	0.201	0.086	0.006	0.000	0.017	0.000
SIWA	0.000	0.000	0.323	0.000	0.000	0.062	0.031	0.187	0.094	0.031	0.021	0.000	0.000
NORTH AF.	0.001	0.000	0.253	0.020	0.001	0.025	0.008	0.215	0.087	0.016	0.005	0.018	0.000
IVOC	0.000	0.000	0.177	0.019	0.007	0.180	0.007	0.050	0.232	0.067	0.011	0.006	0.000
					Haplotypes without the Alu *insertion (*Alu-*)*								
BASQ	0.000	0.000	0.025	0.161	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NSPA	0.000	0.000	0.022	0.303	0.005	0.000	0.000	0.027	0.004	0.000	0.000	0.000	0.000
PASV	0.000	0.000	0.000	0.161	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CSPA	0.000	0.000	0.017	0.255	0.000	0.006	0.011	0.000	0.000	0.000	0.000	0.000	0.000
SSPA	0.000	0.000	0.000	0.293	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000
SFRAN	0.000	0.000	0.011	0.239	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000	0.000
GREEC	0.000	0.000	0.011	0.263	0.000	0.000	0.000	0.006	0.005	0.000	0.000	0.000	0.000
TURK	0.000	0.000	0.007	0.244	0.000	0.000	0.000	0.005	0.004	0.005	0.000	0.000	0.000
GERM	0.000	0.000	0.093	0.285	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EUROPEANS	0.000	0.000	0.016	0.245	0.001	0.001	0.001	0.007	0.002	0.001	0.000	0.000	0.000
ARAB	0.000	0.000	0.067	0.254	0.003	0.000	0.000	0.018	0.007	0.004	0.000	0.000	0.000
NEBE	0.000	0.000	0.008	0.325	0.020	0.009	0.000	0.012	0.005	0.000	0.000	0.000	0.000
MABE	0.000	0.014	0.073	0.250	0.005	0.004	0.000	0.035	0.012	0.003	0.000	0.000	0.000
ASBE	0.000	0.000	0.033	0.276	0.005	0.011	0.000	0.016	0.000	0.000	0.000	0.000	0.005
AMBE	0.000	0.000	0.017	0.226	0.000	0.012	0.000	0.070	0.008	0.000	0.000	0.006	0.000
MZAB	0.000	0.000	0.027	0.270	0.017	0.003	0.000	0.000	0.000	0.005	0.000	0.000	0.000
SIWA	0.000	0.000	0.042	0.146	0.052	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000
NORTH AF.	0.000	0.002	0.039	0.259	0.012	0.006	0.000	0.024	0.005	0.002	0.000	0.001	0.001
IVOC	0.000	0.000	0.067	0.104	0.010	0.020	0.010	0.000	0.034	0.000	0.000	0.000	0.000

			STR variation in Alu+ *chrom.*			*STR variation in* Alu- *chrom.*		
	N	*Haplotype GD values*	Mean	Var	STR GD	Mean	Var	STR GD
Europeans	1436	0.708	96.57	12.92	0.554	90.58	4.4	0.174
North Afr.	1214	0.811	100.73	13.94	0.705	91.80	7.2	0.458
Ivory C.	180	0.858	103.87	12.95	0.778	93.72	9.9	0.721

N: number of chromosomes; GD: Gene Diversity; Mean of the allele size distribution in base pairs; Var: variance of the allele size distribution.

Table S3. FXIIIB *Alu*/STR haplotype frequencies and diversity statistics in 17 populations. STR alleles are expressed in base pairs. Pooled frequencies are also given for the European (9 samples) and North African (7 samples) groups.

	Haplotypes with the Alu *insertion* (Alu+)							*Haplotypes without the* Alu *insertion* (Alu-)						
STR alleles	**172**	**176**	**180**	**184**	**188**	**192**	**196**	**172**	**176**	**180**	**184**	**188**	**192**	**196**
BASQ	0.000	0.010	0.283	0.091	0.131	0.000	0.000	0.078	0.000	0.006	0.066	0.330	0.005	0.000
NSPA	0.021	0.031	0.223	0.066	0.095	0.005	0.000	0.081	0.012	0.025	0.100	0.341	0.000	0.000
PASV	0.000	0.006	0.290	0.076	0.152	0.000	0.000	0.044	0.000	0.026	0.133	0.265	0.006	0.000
CSPA	0.000	0.021	0.231	0.071	0.070	0.000	0.000	0.122	0.011	0.024	0.115	0.334	0.000	0.000
SSPA	0.013	0.000	0.187	0.070	0.103	0.000	0.000	0.129	0.005	0.000	0.128	0.363	0.000	0.000
SFRAN	0.000	0.000	0.211	0.049	0.091	0.000	0.000	0.074	0.011	0.023	0.185	0.356	0.000	0.000
GREEC	0.005	0.026	0.390	0.061	0.086	0.000	0.000	0.036	0.000	0.006	0.090	0.300	0.000	0.000
TURK	0.006	0.005	0.314	0.122	0.138	0.000	0.000	0.068	0.000	0.000	0.091	0.255	0.000	0.000
GERM	0.019	0.000	0.124	0.000	0.034	0.000	0.000	0.077	0.000	0.053	0.306	0.369	0.000	0.016
EUROPEANS	0.007	0.013	0.264	0.074	0.106	0.001	0.000	0.080	0.004	0.014	0.116	0.319	0.001	0.001
ARAB	0.011	0.011	0.121	0.060	0.000	0.000	0.000	0.244	0.021	0.070	0.207	0.255	0.000	0.000
NEBE	0.000	0.000	0.261	0.066	0.057	0.000	0.000	0.243	0.024	0.016	0.187	0.147	0.000	0.000
MABE	0.007	0.000	0.248	0.063	0.074	0.008	0.000	0.231	0.046	0.019	0.174	0.130	0.000	0.000
ASBE	0.013	0.000	0.192	0.074	0.060	0.008	0.000	0.257	0.030	0.005	0.204	0.158	0.000	0.000
AMBE	0.019	0.005	0.167	0.075	0.057	0.000	0.000	0.281	0.095	0.010	0.094	0.189	0.008	0.000
MZAB	0.033	0.014	0.168	0.098	0.064	0.000	0.006	0.250	0.041	0.21	0.163	0.141	0.000	0.000
SIWA	0.000	0.014	0.210	0.057	0.032	0.000	0.000	0.353	0.076	0.014	0.064	0.179	0.000	0.000
NORTH AF.	0.011	0.005	0.205	0.071	0.054	0.003	0.001	0.262	0.046	0.018	0.160	0.162	0.001	0.000
IVOC	0.050	0.052	0.039	0.154	0.016	0.005	0.000	0.375	0.201	0.007	0.082	0.013	0.007	0.000

			STR variation in Alu+ *chrom.*			*STR variation in* Alu- *chrom.*		
	N	*Haplotype GD values*	Mean	Var	STR GD	Mean	Var	STR GD
Europeans	1454	0.7917	182.25	3.69	0.575	184.45	5.7	0.565
North Afr.	1240	0.8286	181.88	3.72	0.582	179.49	6.8	0.697
Ivory C.	174	0.7813	180.61	5.07	0.692	175.18	4.6	0.594

N: number of chromosomes; GD: Gene Diversity; Mean of the allele size distribution in base pairs; Var: variance of the allele size distribution.

Table S4a. DM *Alu*/STR haplotype frequencies and diversity statistics in 17 populations. STR alleles are expressed in base pairs. Pooled frequencies are also given for the European (9 samples) and North African (7 samples) groups.

STR alleles	*Haplotypes with the* Alu *insertion (*Alu+*)*													
	77	**80**	**83**	**86**	**92**	**95**	**98**	**101**	**104**	**107**	**110**	**113**	**116**	**119**
BASQ	0.352	0.005	0.000	0.000	0.024	0.000	0.011	0.005	0.034	0.000	0.000	0.000	0.000	0.000
NSPA	0.295	0.000	0.000	0.000	0.011	0.006	0.007	0.007	0.028	0.006	0.011	0.000	0.000	0.011
PASV	0.399	0.000	0.000	0.000	0.013	0.000	0.029	0.007	0.039	0.000	0.000	0.000	0.000	0.000
CSPA	0.348	0.000	0.000	0.000	0.011	0.000	0.037	0.008	0.045	0.011	0.011	0.000	0.000	0.000
SSPA	0.309	0.000	0.000	0.010	0.000	0.017	0.006	0.000	0.022	0.000	0.005	0.010	0.000	0.005
SFRAN	0.377	0.000	0.000	0.000	0.022	0.000	0.023	0.000	0.011	0.011	0.011	0.011	0.011	0.022
GREEC	0.361	0.000	0.000	0.006	0.006	0.000	0.014	0.015	0.009	0.000	0.011	0.000	0.006	0.000
TURK	0.431	0.000	0.000	0.000	0.010	0.009	0.005	0.008	0.026	0.000	0.005	0.000	0.005	0.010
GERM	0.252	0.000	0.013	0.000	0.000	0.000	0.000	0.015	0.017	0.000	0.000	0.000	0.000	0.000
EUROP.	0.352	0.001	0.001	0.002	0.011	0.004	0.015	0.007	0.027	0.003	0.006	0.002	0.002	0.005
ARAB	0.269	0.000	0.000	0.000	0.000	0.033	0.029	0.023	0.022	0.02	0.000	0.000	0.000	0.011
NEBE	0.286	0.000	0.048	0.000	0.008	0.012	0.007	0.025	0.045	0.011	0.000	0.000	0.000	0.005
MABE	0.282	0.000	0.000	0.000	0.010	0.000	0.005	0.000	0.026	0.000	0.009	0.000	0.000	0.000
ASBE	0.060	0.000	0.008	0.000	0.004	0.009	0.034	0.013	0.025	0.000	0.000	0.000	0.000	0.000
AMBE	0.304	0.000	0.008	0.000	0.008	0.057	0.008	0.008	0.028	0.000	0.000	0.000	0.000	0.000
MZAB	0.242	0.000	0.030	0.000	0.024	0.013	0.019	0.006	0.036	0.000	0.000	0.000	0.000	0.000
SIWA	0.189	0.000	0.000	0.000	0.078	0.015	0.000	0.016	0.010	0.000	0.014	0.000	0.007	0.000
N. AF.	0.283	0.000	0.014	0.000	0.017	0.016	0.015	0.012	0.028	0.003	0.003	0.000	0.001	0.002
IVOC	0.159	0.000	0.023	0.000	0.037	0.030	0.027	0.039	0.158	0.000	0.043	0.006	0.000	0.000

	122	**125**	**128**	**131**	**134**	**137**	**140**	**143**	**146**	**149**	**152**	**155**	**158**
BASQ	0.015	0.010	0.015	0.010	0.005	0.005	0.020	0.000	0..010	0.000	0.005	0.020	0.000
NSPA	0.000	0.023	0.017	0.011	0.011	0.006	0.006	0.000	0.011	0.000	0.000	0.000	0.000
PASV	0.033	0.000	0.027	0.000	0.000	0.000	0.033	0.007	0.000	0.000	0.000	0.000	0.000
CSPA	0.017	0.022	0.022	0.006	0.006	0.011	0.000	0.011	0.006	0.011	0.000	0.000	0.000
SSPA	0.005	0.019	0.016	0.016	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000
SFRAN	0.000	0.044	0.022	0.011	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000
GREEC	0.017	0.039	0.028	0.006	0.011	0.000	0.011	0.000	0.006	0.000	0.000	0.000	0.000
TURK	0.010	0.016	0.010	0.010	0.005	0.005	0.000	0.000	0.000	0.000	0.000	0.005	0.005
GERM	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EUROP.	0.012	0.019	0.018	0.008	0.005	0.003	0.009	0.003	0.004	0.001	0.001	0.003	0.001
ARAB	0.011	0.011	0.023	0.011	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000
NEBE	0.005	0.016	0.000	0.005	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000
MABE	0.009	0.017	0.013	0.004	0.009	0.000	0.000	0.004	0.000	0.000	0.004	0.000	0.000
ASBE	0.004	0.025	0.008	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AMBE	0.000	0.027	0.008	0.007	0.000	0.014	0.007	0.000	0.000	0.000	0.000	0.000	0.000
MZAB	0.000	0.000	0.006	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SIWA	0.007	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N. AF.	0.005	0.015	0.010	0.006	0.002	0.002	0.002	0.002	0.000	0.000	0.001	0.000	0.000
IVOC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

N: number of chromosomes; GD: Gene Diversity; Mean of the allele size distribution in base pairs; Var: variance of the allele size distribution.

Table S4b. DM *Alu*/STR haplotype frequencies and diversity statistics in 17 populations. STR alleles are expressed in base pairs. Pooled frequencies are also given for the European (9 samples) and North African (7 samples) groups.

	Haplotypes without the Alu *insertion* (Alu-)										
	65	**77**	**80**	**83**	**86**	**89**	**92**	**95**	**98**	**101**	**104**
BASQ	0.005	0.011	0.000	0.000	0.000	0.000	0.000	0.093	0.166	0.137	0.029
NSPA	0.000	0.030	0.006	0.000	0.000	0.000	0.000	0.118	0.123	0.185	0.040
PASV	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.033	0.171	0.160	0.028
CSPA	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.112	0.0121	0.149	0.006
SSPA	0.000	0.064	0.000	0.000	0.010	0.000	0.000	0.057	0.225	0.168	0.026
SFRAN	0.000	0.034	0.000	0.000	0.000	0.000	0.000	0.089	0.099	0.133	0.033
GREEC	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.0128	0.152	0.134	0.019
TURK	0.000	0.011	0.005	0.000	0.000	0.000	0.000	0.107	0.105	0.150	0.016
GERM	0.000	0.113	0.000	0.013	0.000	0.000	0.013	0.108	0.230	0.187	0.024
EUROP.	0.001	0.025	0.001	0.001	0.003	0.000	0.001	0.094	0.153	0.155	0.024
ARAB	0.000	0.049	0.000	0.000	0.000	0.000	0.000	0.092	0.232	0.079	0.046
NEBE	0.000	0.006	0.000	0.000	0.000	0.011	0.003	0.084	0.259	0.087	0.014
MABE	0.000	0.049	0.000	0.000	0.013	0.000	0.002	0.159	0.224	0.112	0.017
ASBE	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.123	0.228	0.080	0.005
AMBE	0.000	0.045	0.000	0.000	0.000	0.000	0.000	0.149	0.204	0.061	0.013
MZAB	0.000	0.049	0.000	0.012	0.000	0.000	0.000	0.190	0.308	0.017	0.006
SIWA	0.000	0.090	0.000	0.000	0.000	0.000	0.007	0.049	0.257	0.069	0.112
N. AF.	0.000	0.039	0.000	0.002	0.002	0.002	0.002	0.125	0.244	0.075	0.025
IVOC	0.000	0.029	0.000	0.000	0.000	0.000	0.039	0.070	0.237	0.067	0.024
	107	**110**	**113**	**122**	**125**	**128**	**131**	**137**	**140**	**143**	**152**
BASQ	0.000	0.000	0.000	0.000	0.005	0.000	0.005	0.000	0.000	0.000	0.005
NSPA	0.006	0.000	0.000	0.006	0.010	0.006	0.006	0.000	0.000	0.000	0.000
PASV	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000
CSPA	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SSPA	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000
SFRAN	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000
GREEC	0.011	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TURK	0.010	0.005	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GERM	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EUROP.	0.007	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.000	0.001
ARAB	0.014	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000
NEBE	0.048	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MABE	0.017	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.004	0.000
ASBE	0.055	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AMBE	0.048	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MZAB	0.024	0.006	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000
SIWA	0.021	0.007	0.014	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000
N. AF.	0.034	0.003	0.002	0.000	0.002	0.003	0.001	0.000	0.000	0.001	0.000
IVOC	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

			STR variation in Alu+ *chrom.*			*STR variation in* Alu- *chrom.*		
	N	Haplotype GD values	Mean	Var	STR GD	Mean	Var	STR GD
Europeans	1434	0.8166	90.33	21.38	0.5145	98.11	7.4	0.7305
North Afr.	1198	0.8330	87.35	16.86	0.5695	97.41	7.4	0.8276
Ivory C.	170	0.8739	93.54	12.39	0.7906	96.75	6.0	0.7014

N: number of chromosomes; GD: Gene Diversity; Mean of the allele size distribution in base pairs; Var: variance of the allele size distribution

11 The Ins and Outs of Population Relationships in West-Mediterranean Islands: Data from Autosomal Alu polymorphisms and Alu/STR compound systems

Emili González-Pérez, Pedro Moral, Marc Via, Giuseppe Vona, Laurent Varesi, Josep Santamaria, Magdalena Gaya-Vidal, Esther Esteban.

Journal of Human Genetics 2007; 52(12): 999-1010

11.1 Relacions poblacionals a les illes del Mediterrani occidental: dades de polimorfismes Alu i sistemes compostos Alu/STR

Aquest treball continua el plantejament metodològic expressat en l'article anterior inclòs en aquesta tesi doctoral. La metodologia de base és la determinació de la variació de dues bateries de marcadors de naturalesa genètica diferent i que per tant poden aportar explicacions complementàries a l'hora de la reconstrucció poblacional de les poblacions humanes: una bateria de 18 elements *Alu* autosòmics y l'element específic del cromosoma Y (YAP) i els tres sistemes autosòmics compostos *Alu*/STR (dels *loci* CD4, FXIIIB i DM).

L'estudi ha estat aplicat, en aquesta ocasió i de manera més concreta, a la descripció de la variació genètica d'aquests marcadors en les grans illes del Mediterrani occidental i en la seva interpretació dins del marc evolutiu de les poblacions humanes específiques de la conca mediterrània.

Les illes del Mediterrani occidental han representat des d'antic un veritable pal de paller molt destacat a l'hora de comprendre adequadament els estudis arqueològics, històrics i antropològics duts a terme, com a conseqüència d'un paper central i actiu en el desenvolupament i història de les principals civilitzacions mediterrànies. Tot i això, les dades genètiques disponibles encara són escasses, especialment a l'hora d'interpretar el seu grau de diferenciació interna i d'establir el tipus de relacions poblacionals humanes que presenten amb la resta de grups humans de la regió.

Així, i més concretament, les poblacions estrictament analitzades per aquest estudi, han estat un total de set: Mallorca, Còrsega interior, litoral occidental de Còrsega, Sardenya interior, litoral occidental de Sardenya, Sicília occidental i Sicília oriental. La base de dades obtinguda s'ha completat amb les dades complementàries disponibles per a les poblacions continentals mediterrànies amb l'objectiu de deduir i interpretar les relacions humanes globals esmentades a la regió.

L'anàlisi genètic d'aquestes poblacions insulars mostra, especialment per a la bateria de 18 elements *Alu* polimòrfics estudiats, un grau considerablement elevat d'heterogeneïtat genètica tant interna com entre les diferents illes. A més a més, cal destacar que la diferenciació global entre illes (Fst 2,2 %) és, fins i tot, superior a l'observada per al conjunt de poblacions continentals mediterrànies (europees i nord-africanes) pels mateixos marcadors.

Les dades obtingudes a partir dels càlculs de divergències poblacionals i l'elevada heterogeneïtat suggereixen que el patró genètic actual d'aquestes poblacions insulars conserva característiques diferenciadores que es remuntarien als primers pobladors humans de les mateixes, en el període paleolític.

D'altra banda, i en contraposició amb la informació aportada exclusivament pels elements *Alu*, la variació mostrada pels marcadors STR i pels haplotips *Alu*/STR –marcadors amb taxes de mutació superiors i més afectats per les vicissituds poblacionals recents– indica un grau d'homogeneïtat més marcada, suggerint que, com a mínim des del neolític, el flux gènic ha estat actuant força contínuament en tot el Mediterrani occidental.

Finalment, les dades que obtenim dels resultats poden ser indicatius també de certs aspectes particulars. En aquest sentit, caldria destacar un efecte important de la deriva genètica i l'aïllament poblacional des de ben antic a Sardenya, un grau de flux gènic nord-africà considerable representat per diferents haplotips típicament berbers a Sicília (especialment a la part més occidental) i unes intenses relacions històriques recents de Còrsega, Mallorca i la Sicília més oriental amb les poblacions humanes continentals de la riba mediterrània europea.

11.2 Informe del director sobre la participació del doctorand

El Dr. **Pedro Moral Castrillo**, Professor Titular del Departament de Biologia Animal de la Universitat de Barcelona i director de la tesi doctoral **"Variació genètica i evolució d'elements *Alu* recents en poblacions humanes. Inferències biodemo-gràfiques i filogeogràfiques"** presentada pel doctorand **Emili González Pérez**, desitja fer constar que la participació del doctorand en l'elaboració de l'article **"The Ins and Outs of Population Relationships in West-Mediterranean Islands: Data from Autosomal *Alu* polymorphisms and *Alu*/STR Compound Systems"** publicada per la revista *Journal of Human Genetics*, ha consistit en les següents tasques principals:

- Processament de mostres: extracció de DNA de les mostres poblacionals utilitzades en el treball (100 %)
- Disseny del treball i selecció de marcadors analitzats, conjuntament amb el Dr. Pedro Moral
- Genotipatge d'insercions *Alu* i STRs / Seqüenciacions (100 %)
- Elaboració de les bases de dades de resultats (100 %)
- Anàlisi estadística dels resultats, conjuntament amb la Dra. Esther Esteban
- Redacció de l'article (amb la Dra. Esteban i el Dr. Moral)

Complementàriament, cal assenyalar que cap dels coautors d'aquest article ha utilitzat, ni implícitament ni explícitament, els resultats d'aquest treball per a l'elaboració d'una altra tesi doctoral. En conseqüència, aquest treball forma part, en exclusiva, del treball de recerca en el que s'emmarca la tesi del doctorand Emili González Pérez.

Signat: Dr. Pedro Moral Castrillo
Barcelona, 29 de setembre de 2009

J Hum Genet (2007) 52:999–1010
DOI 10.1007/s10038-007-0206-6

ORIGINAL ARTICLE

The ins and outs of population relationships in west-Mediterranean islands: data from autosomal *Alu* polymorphisms and *Alu*/STR compound systems

E. Gonzalez-Perez · P. Moral · M. Via · G. Vona · L. Varesi · J. Santamaria · M. Gaya-Vidal · E. Esteban

Received: 25 July 2007 / Accepted: 28 September 2007 / Published online: 24 October 2007
© The Japan Society of Human Genetics and Springer 2007

Abstract The islands of the West Mediterranean have played a central role in numerous archaeological, historical and anthropological studies due to their active participation in the history of main Mediterranean civilisations. However, genetic data failed to fit in both their degree of internal differentiation and relationships. A set of 18 *Alu* markers and three short tandem repeats (STRs) closely linked to the CD4, F13B and DM *Alu* have been analysed in seven samples from Majorca, Corsica, Sardinia and Sicily to explore some of these issues. Our samples show a high genetic heterogeneity inside and among islands for the *Alu* data. Global differentiation among islands (F_{ST} 2.2%) is slightly higher than that described for Europeans and North Africans. Both the estimated divergence times among samples and the high population heterogeneity revealed by *Alu* data are compatible with population differences since the first islands' settlement in the Paleolithic period. However, the high within-population diversities and the remarkable homogeneity observed in both STR and *Alu*/STR haplotype variation indicated that, at least since Neolithic times, gene flow has been acting in west Mediterranean. Genetic drift in west-coast Sardinia and gene flow in west Sicily have contributed to their general differentiation, whereas Corsica, Majorca and east Sicily seem to reflect more recent historical relationships from continental south Europe.

Keywords *Alu* insertions · *Alu*/STR haplotypes · Human populations · Mediterranean peopling

Electronic supplementary material The online version of this article (doi:10.1007/s10038-007-0206-6) contains supplementary material, which is available to authorised users.

E. Gonzalez-Perez · P. Moral · M. Via · J. Santamaria · M. Gaya-Vidal · E. Esteban
Unitat d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Barcelona, Spain

G. Vona
Dipartamento di Biologia Sperimentale, Università di Cagliari, Cagliari, Italy

L. Varesi
Department of Human Genetics, Faculty of Science, University of Corsica, Corte, France

E. Esteban (✉)
Unitat d'Antropologia, Facultat de Biologia, Avda. Diagonal 645, 08028 Barcelona, Spain
e-mail: mesteban@ub.edu

Introduction

The Balearic archipelago and the islands of Sardinia, Corsica and Sicily are enclosed in the westernmost part of the Mediterranean basin by the Iberian Peninsula and the Strait of Gibraltar to the west, North Africa to the south and the Strait of Messina and the Italic Peninsula to the east. This region covers an area of about 0.85 million km^2 that embraces a set of populations closely related not only by geography but also by historical relationships, probably since their initial peopling in middle and upper Paleolithic times. Their common historical background includes numerous and almost continuous waves of settlements and conquests by several mainland civilisations. This coming and going of populations in a close geographical area has constituted a challenge for archaeologists, historians, ethnologists and anthropologists alike. The latter have been particularly interested in determining the degree of genetic relationships of a set of populations that, even having a

common past of invaders for centuries, still preserve some remarkable differences.

In Majorca, the largest island of the Balearic archipelago, archaeological data suggest peopling since the Paleolithic period. The island was occupied by the Carthaginians before passing to the Romans, who installed a long period of prosperity. From 707, the island was increasingly attacked by Muslim raiders from North Africa. Two centuries later, the Caliphate of Cordoba conquered Majorca, ushering in a new period of prosperity for the island. In the thirteen century, the Catalano–Aragonese launched an invasion with 15,000 men and 1,500 horses, annexing the island to the kingdom. In the archipelago, the mother tongue is the Balearic variation of Catalan, a Romance language spoken in a large part of the former territories of this kingdom. From a genetic point of view, recent data from mtDNA haplotype variability (Picornell et al. 2005) suggest a high similarity among Majorca, other Balearic islands and Spanish populations historically related with the Catalano–Aragonese kingdom. This affinity points to an important gene flow from the mainland without significant bottlenecks involved in the colonisation of the island.

Corsica and Sardinia formed a single land mass in early Paleolithic times. They are now separated by a straight of about 12 km wide. As a result of this close vicinity, these two islands share a common background despite the fact that for many centuries, the contact with different Mediterranean invaders was apparently limited to the coastal flatland territories of the islands. Carthaginians and Romans pushed the indigenous people into the central region of the islands, which explains the fact that, in the case of Sardinia, the centre is the most conservative region linguistically and genetically (Piazza et al. 1988; Cappello et al. 1996). In regards to the vernacular languages, the inhabitants of both islands speak Romance languages, but in the case of Corsica, the language shows a high affinity with the Tuscan dialect although with some internal differentiations, whereas Sardinian is a clearly distinct Romance language, preserving traces of the indigenous pre-Roman languages of the island until this very day.

These two islands also share some demographic features, likely a product of their abrupt geography, that have modeled their genetic structure mainly due to the effect of isolation and genetic drifting. Until the eighteenth century, Sardinia had a population that rarely exceeded 400,000 inhabitants. It was even lower in Corsica: 100,000 inhabitants until the end of the eighteenth century (Day 1987; Gatti 1995; Simi 1997). Genetic studies conducted in Corsica and Sardinia, although numerous, failed to coincide in data. Some studies based on classical markers indicate genetic similarity (Memmi et al. 1998; Vona et al. 2003), whereas others emphasise their genetic heterogeneity (Calafell et al. 1996). More recently, DNA studies have added more data without conclusive results. Francalacci et al. (2003) conducted a survey of Y-chromosome haplotypes in several samples from Corsica, Sicily and central Sardinia. Their main findings underline the differentiation of Sardinians and Sicilians from other Mediterraneans, whereas Corsica remains more similar to continental Italy and French samples, excluding the possibility of significant gene flow from central Sardinia to north-central Corsica. Data from mtDNA (Morelli et al. 2000) have also demonstrated a remarkable discontinuity among central Sardinians and both north Sardinians and Corsicans. On the other hand, a different mtDNA study (Falchi et al. 2006) found genetic similarities among Iberian, Corsican and Sardinian populations. This study confirms the fact that most mtDNA haplogroups in these samples coalesced in Paleolithic dates. Information from autosomal markers also gives controversial results; the maternal genetic similarities among Iberian, Corsican and Sardinian populations seem to be reflected in the high frequencies of ß039 thalassemic mutation (Falchi et al. 2005), whereas a multilocus analysis of autosomal microsatellites (Tofanelli et al. 2001) suggests a remarkable genetic differentiation between Sardinia and Corsica.

The particular position of Sicily in the centre of the Mediterranean has made the passage through it easier for peoples from virtually all of the Mediterranean and beyond. Before the Roman conquest, Sicily was occupied by remnants of the autochthonous populations of Sicani, Elymi, and Siculi (Indo-European populations that arrived between the second and first millennium BC), as well as by Phoenicians (tenth to eighth century BC) and Greeks (eighth century BC). The Sicilian language has inherited vocabulary and grammatical forms from these earliest settlers of the island as well as from the later colonists and conquerors. In view of their heterogeneous background, the subject of genetic relationships between populations on the island of Sicily is controversial. Some studies based on classical polymorphisms, and later on autosomal DNA markers (Calò et al. 2003; Ghiani et al. 2002; Piazza et al. 1988; Romano et al. 2003), indicated that Sicily is genetically heterogeneous, with a considerable East–West gradient compatible with population settlements occurring at different times. Other authors (Rickards et al. 1998) state that there was no clear geographic clustering within Sicily, rejecting an East–West differentiation.

Although the genetic information here summarised is extensive and covers everything from classical polymorphisms to uniparental and autosomal DNA, as far as we know, none of these studies have tested the four main islands with samples including different geographical areas inside each one jointly. This is the context in which we are presenting our work. We analysed a set of eight autosomal *Alu* polymorphisms and three short tandem

repeats (STRs) closely linked to the CD4, F13B and DM *Alu* markers in seven regions of Majorca, Sardinia, Corsica, and Sicily. We selected these particular markers for two main reasons, the first being the widely contrasted informative nature of *Alu* insertions for the study of human populations (Watkins et al. 2001) due to their stability, low mutation rate and known ancestral state, and the second due to the remarkable degree of information provided by *Alu* markers linked with STRs. The latter are very effective for estimating divergence between populations, although their mutation rate involves a certain degree of homoplasy that can mask the true genetic relationships. Information on haplotype frequencies, together with STR variation on ancestral *Alu* allele background compared with STR variation on the derived *Alu* alleles, has been used to estimate fine genetic relationships between human populations, not only on a large geographical scale (Tishkoff et al. 1996; Ramakrishnan and Mountain 2004), but also at a microgeographical level (Flores et al. 2000; Esteban et al. 2004).

The main objectives of this work are: (1) exploration of the degree of internal variability of Corsica, Sardinia and Sicily for comparison of the results with previous studies that suggested different heterogeneity levels inside these islands, (2) analysis of the genetic relationships among the four islands by maximum usage of different genetic markers such as *Alu* and STRs and (3) use of the qualitative information provided by the *Alu*/STR haplotypes to determine the amount of external gene flow received in the islands as a result of their historical background.

Material and methods

A total of 360 unrelated and healthy autochthonous individuals from seven well-defined rural areas of Majorca, central Sardinia, west-coast Sardinia, central Corsica, west-coast Corsica, east Sicily and west Sicily were analysed. Samples were obtained with the informed consent of the participants, whos four grandparents were born in the same region. The geographical position of the samples is detailed in Fig. 1.

Eight human-specific *Alu* insertion polymorphisms (DM, HS2.43, B65, PV92, D1, F13B, A25 and TPA25) were typed using the primers and polymerase chain reaction (PCR) amplification conditions previously described in Stoneking et al. (1997) and Edward and Gibbs (1992), with minor modifications. As for STRs, CD4 consists of a pentanucleotide $(TTTTC)_n$ repeat amplified according to Tishkoff et al. (1996), with minor modifications. This STR maps approximately 9 kb from the *Alu* marker. The F13B STR is a tetranucleotide repeat $(TTTA)_n$ at 4 kb of the *Alu* marker. Amplification conditions were as described in Nishimura and Murray (1992), with slight modifications. The DM $(CTG)_n$ repeat was amplified according to Brook et al. (1992). In this case the *Alu* polymorphism is located 5 kb telomeric to the repeat. After amplification with fluorescent-labeled primers, PCR products were pooled and electrophoresed on an ABI PRISM 3700 DNA sequencer (Applied Biosystems, Foster City, CA, USA). Genescan and Genemapper 3.0 programs (ABI PRISM, Applied Biosystems) were used to generate fragment sizes and

Fig. 1 Geographical position of the seven insular samples and other west Mediterranean groups used in comparisons

genotypes. Different selected individuals were sequenced for each STR to confirm size lengths and assign the correct repeat number for comparisons with data generated by other authors.

Allele frequencies were computed by direct counting, and Hardy–Weinberg equilibrium was tested by an exact test (Guo and Thomson 1992). Standard gene diversity indices by populations and locus were estimated according to Nei (1987). Locus frequency distributions were compared by an exact test for population differentiation. The program PHASE was used to generate, by means of a Bayesian statistical method, estimates of haplotype frequencies. The v.2.1 implements extensions of the original methods described in Stephens et al. (2001) and Stephens and Donnelly (2003). Linkage disequilibrium estimates between the STR and their respective *Alu* were quantified using the adaptation of Black and Krafsur (1985) algorithms contained in the computer program GENETIX 4.05 (Belkhir et al. 1996–2004). The apportionment of genetic variance was checked by analysis of molecular variance (AMOVA) through the ARLEQUIN computer package (Excoffier et al. 2005). Locus-by-locus fixation indices (F_{ST}, F_{SC} and F_{CT}) were averaged to obtain a global value. The statistical significance of these averages was checked by combining probabilities (Sokal and Rohlf 1997).

Apart from the seven samples included in this study, data from other European and North African samples, mainly from the Mediterranean basin, were collected from the literature. For our seven samples, we reached a database of 18 *Alu* polymorphisms by linking this work to a previous one conducted by our research team (Caló et al. 2005), but the available literature allowed us to create a database of 20 Mediterranean groups only for the following *Alu*: APO, B65, PV92, D1, F13B, A25, TPA25 and ACE. These samples (see Fig. 1 for geographical location) come from the works of Stoneking et al. (1997), Comas et al. (2000) and Garcia-Obregon et al. (2006, 2007). The consulted population data from CD4, F13B and DM STRs was obtained from the ALFRED database (Rajeevan et al. 2005).

Population relationships were approached by means of F_{ST}-related genetic distances analyses (Reynolds et al. 1983) using the PHYLIP 3.6 package (Felsenstein 1989) and depicted through multidimensional scaling from the distance matrix. Genetic distances $(\delta\mu)^2$ for STR data according to Goldstein et al. (1995) were calculated by the computer program Microsat 2 (written by E. Minch and available from: http://www.hpgl.stanford.edu). Population divergence times were estimated according to Goldstein et al. (1995), who proposed an equation to calculate divergence time among two samples by dividing the estimated value of the $(\delta\mu)^2$ distance by twice the product of the mutation rate (β) and the constant size variance (ω) of mutational jumps considering a generation time of 25 years. For divergence time calculations, we assumed that ω is constant with a value of 0.04 (1/25) and β value of 2.8 (10^{-4} (Chakraborty et al. 1997). The time obtained is expressed in years before present (YBP).

Results

Variability in west Mediterranean Islands

Alu *polymorphisms*

Allele frequencies for the eight *Alu*s are shown in Table 1. In general, all loci were in Hardy–Weinberg equilibrium after Bonferroni correction (excepting D1 in west Sicily) and showed significant gene diversity differences (Kruskal–Wallis test $p < 0.001$). In regards to F_{ST} values, only the PV92 loci showed moderate genetic differentiation (F_{ST} of 6.7%). Although the samples from Sicily and Sardinia occupied extreme positions in the population variation ranges (see Table 1) for some loci, when average heterozygosities were compared, the Kruskal–Wallis test indicated no remarkable differences ($p = 0.873$) among our samples.

Pairwise population comparisons across the eight loci revealed a remarkable degree of heterogeneity (significant p values for 17 comparisons out of 21) disrupted only for four population comparisons that failed to show significant differences: west-coast Corsica with both central Sardinia and Majorca, and central Corsica with east Sicily and Majorca. The locus that showed the highest number of significant population comparisons (11 out of 21) was consistent with the genetic differentiation revealed by F_{ST} values, PV92. A nonhierarchical AMOVA yielded an average F_{ST} in west-Mediterranean islands of 2.2% ($p < 0.001$). Inside Sardinia, the level of population genetic variance (F_{ST} of 5.5%, $p < 0.001$) was even higher than that observed for the whole of the islands. Both the low number of samples and the extreme allele frequencies shown by the two Sardinian samples in almost all loci probably accounted for this F_{ST} value. However, the same pattern was observed when we recalculated F_{ST} using information from 18 *Alu* polymorphisms: average F_{ST} (3.7%, $p < 0.001$) inside Sardinia was triple that observed among islands (1.3%, $p < 0.01$).

STR gene diversity and Alu*/STR compound systems*

Allele size frequencies of CD4, F13B and DM microsatellites are available as supplementary material from the Web site of the journal. Overall, the three distributions were in Hardy–Weinberg equilibrium after Bonferroni correction, with the only exception of DM STR being in central

Table 1 *Alu* insertion frequencies in west-Mediterranean islands

Alu	Sardinia centre	Sardinia west coast	Corsica centre	Corsica west coast	Sicily east	Sicily west	Majorca	F_{ST} per locus	Variation ranges reviewed populations
DM (2N)	100	74	92	90	80	92	110		Loci not included in the reviewed literature
Alu+	0.607	0.300	0.695	0.567	0.787	0.576	0.527	0.030	
H	0.485	0.442	0.428	0.497	0.339	0.494	0.503	$p = 0.043$	
HS2.43 (2N)	100	98	102	76	102	100	114		0.02 SE Morocco – 0.28 C Sardinia
Alu+	0.280	0.061	0.137	0.158	0.069	0.130	0.096	0.030	
H	0.407	0.116	0.239	0.269	0.129	0.228	0.176	$p = 0.020$	
B65 (2N)	100	46	102	94	102	96	106		0.47 Navarre (Spain) – 0.73 Algerians
Alu+	0.520	0.630	0.598	0.489	0.569	0.583	0.575	0	
H	0.504	0.476	0.485	0.505	0.485	0.491	0.493	NS	
PV92 (2N)	98	90	102	94	98	90	112		0.04 W Sicily – 0.40 S Morocco
Alu+	0.122	0.389	0.245	0.138	0.112	0.044	0.087	0.067	
H	0.217	0.481	0.374	0.241	0.201	0.086	0.307	$p = 0.009$	
D1 (2N)	100	46	102	94	102	96	114		0.11 WC Sardinia – 0.53 W Sicily
Alu+	0.310	0.109	0.323	0.415	0.422	0.531	0.342	0.029	
H	0.432	0.198	0.442	0.491	0.492	0.503	0.454	$p = 0.031$	
F13B (2N)	100	84	102	96	102	94	114		0.29 W Morocco – 0.62 Greek Cypriots
Alu+	0.370	0.333	0.451	0.583	0.471	0.372	0.500	0.009	
H	0.471	0.450	0.500	0.491	0.503	0.472	0.504	NS	
A25 (2N)	100	92	102	94	102	94	114		0.06 C Sardinia – 0.23 SE Morocco
Alu+	0.060	0.109	0.137	0.096	0.108	0.202	0.114	0	
H	0.114	0.196	0.239	0.175	0.194	0.326	0.204	NS	
TPA25 (2N)	100	66	96	94	86	92	114		0.33 WC Sardinia – 0.64 W Sicily
Alu+	0.500	0.333	0.542	0.596	0.465	0.641	0.570	0.012	
H	0.509	0.451	0.502	0.487	0.504	0.465	0.494	$p = 0.052$	
Average H	0.392	0.351	0.401	0.394	0.356	0.383	0.392		Among Islands average F_{ST} 0.022
SD	0.147	0.153	0.109	0.140	0.161	0.155	0.141		$p < 0.001$
CD4									Data from Calo et al. (2005)
Alu+	0.694	0.786	0.716	0.837	0.598	0.700	0.693		

Previously described CD4 *Alu* frequencies on these samples have been included for posterior use in CD4 *Alu*/STR haplotype calculations (Table 3). Variation ranges according to data from the reviewed literature for 20 European and North African populations

2N number of chromosomes analysed, *H* Nei's nonbiased gene diversity *NS* not significant

Sardinia. Allele diversity values and some statistical parameters of allele size distributions, including STR variation on derived chromosomes, are reported in Table 2. Heterozygosity values in the three STRs showed similarly notable levels of within-population variation, but no significant population differences were detected in neither diversity values or allele size frequencies. STR variation in the CD4- and DM-derived chromosomes (those *Alu*-) was extremely lower in all cases, according to previous knowledge about the distribution of these compound systems in modern humans (Tishkoff et al. 1996, 1998). On the contrary, STR variation in the derived F13B chromosomes (those carrying the *Alu* insertion) was high and very similar to that described for the general variation of this STR.

Alu/STR haplotype frequencies are reported in Tables 3, 4 and 5 for CD4, F13B and DM markers, respectively. In the three compound systems, *Alu* and STR alleles were in linkage disequilibrium. Agreeing with that observed for the STR distributions, our samples showed high levels of within-population diversity but weak population differences.

The number of different CD4 haplotypes in Sardinia and Sicily (seven and nine, respectively) exceeded in number those found in Corsica and Majorca due to the presence of some African characteristic combinations (*Alu*—with both alleles of five and eight repeats) in the former populations. In the particular case of east Sicily, these haplotypes accounted for a frequency of 5.6%. Another haplotype (*Alu*-/10 repeats allele) characteristic of Berber groups (Flores et al. 2000;

Table 2 Variation of CD4, F13 and DM microsatellites in west-Mediterranean islands

	Global STR variation						STR variation on derived *Alu* chromosomes					
	2N	No. alleles	Mean	Variance	Allele range	H	2N	No. alleles	Mean	Variance	Allele range	H
	CD4 pentanucleotide											
Sardinia C	90	4	6.75	4.70	5–10 (5)	0.677	27	3	6.11	0.64	5–10 (6)	0.140
Sardinia WC	94	6	7.61	6.14	5–12 (5)	0.724	26	1	6.00	–	6	0.000
Corsica C	94	5	6.91	5.00	5–11 (5)	0.718	31	2	5.93	0.06	5–6 (6)	0.121
Corsica WC	96	6	7.12	5.77	5–12 (5)	0.683	19	1	6.00	–	6	0.000
Sicily east	94	6	6.84	4.61	5–12 (6)	0.704	36	3	6.14	0.98	5–10 (6)	0.248
Sicily west	96	6	7.57	5.73	5–12 (10)	0.724	28	3	6.53	2.11	5–10 (6)	0.304
Majorca	100	4	6.83	5.06	5–11 (5)	0.684	26	1	6.00	–	6	0.000
	F13B tetranucleotide											
Sardinia C	72	4	8.74	1.63	6–10 (10)	0.720	10	2	8.57	0.70	9–10 (10)	0.520
Sardinia WC	76	4	8.84	1.36	6–10 (9)	0.713	27	3	8.88	0.64	8–10 (8/9)	0.658
Corsica C	94	5	8.80	1.54	6–10 (10)	0.708	41	4	8.32	0.57	7–10 (8)	0.468
Corsica WC	96	4	8.64	1.39	6–10 (8)	0.680	56	3	8.46	0.54	8–10 (8)	0.487
Sicily east	88	5	8.58	1.69	6–10 (8)	0.729	41	4	8.44	0.80	7–10 (8)	0.570
Sicily west	92	5	8.73	1.54	6–10 (10)	0.685	35	4	8.74	0.96	7–10 (8)	0.580
Majorca	94	5	8.74	1.68	6–11 (10)	0.734	44	3	8.54	0.53	8–10 (8)	0.558
	DM trinucleotide											
Sardinia C	94	13	9.75	26.95	5–31 (5)	0.760	32	4	9.68	12.16	5–13 (12)	0.101
Sardinia WC	94	15	10.91	33.15	5–29 (5)	0.799	67	6	12.13	6.33	5–15 (12)	0.015
Corsica C	96	12	9.43	35.13	5–33 (5)	0.704	28	6	12.57	2.18	8–15 (13)	0.014
Corsica WC	96	13	10.02	33.37	5–29 (5)	0.732	41	6	11.51	10.81	5–20 (13)	0.083
Sicily east	94	14	10.82	38.77	5–26 (5)	0.784	19	3	12.05	0.61	11–13 (12)	0.157
Sicily west	96	13	10.02	30.41	5–27 (5)	0.762	35	7	12.43	6.37	5–21 (13)	0.036
Majorca	114	14	11.05	34.27	5–30 (5)	0.796	48	6	10.89	12.61	5–21 (13)	0.063

Allele range in number of repeats. In parentheses: modal allele

2N number of chromosomes, *No. alleles* number of different alleles, variance in repeat number, *H* Nei's nonbiased gene diversity

Esteban et al. 2004) was also found in central Sardinia (1.2%), east Sicily (2.5%) and, with more remarkable frequencies, in west Sicily (4%). The pattern of F13B haplotype frequencies was very similar among islands excluding the comparison between west-coast Sardinia and west-coast Corsica ($p = 0.04$). Some haplotypes were common to all groups, whereas some others were found scattered in certain samples; however, none of these particular haplotypes were detected in Majorca. DM haplotypes also showed a similar pattern of high within-population diversity combined with great population homogeneity.

Leaving aside some occasional differences, the remarkable genetic heterogeneity within and among islands detected for the set of *Alu* markers did not match up with the global homogeneity detected for STR variation and *Alu*/STR haplotype frequencies. Furthermore, in west-Mediterranean islands, the F_{ST} values deduced from STR variation in the three loci as a whole ($F_{ST} = 0.01\%$) or from the three *Alu*/STR compound systems ($F_{ST} = 0.02\%$) were not significantly different from zero.

Genetic relationships in the west-Mediterranean basin

Global relationships in our samples were assessed through F_{ST}-related genetic distance matrices for 18 *Alu* polymorphisms and three *Alu*/STR combinations and through $(\delta\mu)^2$ distances for the three STRs. In all matrices, distance values were significantly different from zero in more than 90% of cases. *Alu* and *Alu*/STR distance matrices were positively correlated (Mantel test, $r = 0.763$, $p = 0.041$) and underlined the genetic differentiation of west-coast Sardinia and west Sicily [see Fig. 2a for the multidimensional scaling (MDS) plot based on *Alu* data]. The first dimension of the MDS plot based on $(\delta\mu)^2$ distances (Fig. 2b) clearly distinguished two population clusters, with west-coast Sardinia and west Sicily as the most differentiated samples within each group. Figure 2b also contains estimates of divergence times among samples; the two main population clusters showed a time separation of around 25,000 YBP, whereas the divergence inside each group was considerably lower.

Table 3 CD4 *Alu*/short tandem repeat (STR) haplotype frequencies and global gene diversity

	Haplotypes with the *Alu* insertion						
	5 (85)	6 (90)	8 (100)	9 (105)	10 (110)	11 (115)	12 (120)
Sardinia C	0.3736 ± 0.0042	0.0593 ± 0.0059			0.2375 ± 0.0038		0.0227 ± 0.0002
Sardinia WC	0.3022 ± 0.0023	0.0286 ± 0.0053		0.0132 ± 0.0001	0.3405 ± 0.0043	0.0524 ± 0.0017	0.0394 ± 0.0001
Corsica C	0.3466 ± 0.0008	0.0123 ± 0.0045			0.2518 ± 0.0052	0.0488 ± 0.0062	
Corsica WC	0.4355 ± 0.0047			0.0104 ± 0.0001	0.3008 ± 0.0036	0.0417 ± 0.0001	0.0104 ± 0.0001
Sicily east	0.2946 ± 0.0059	0.0500 ± 0.0094			0.1875 ± 0.0074	0.0423 ± 0.0017	0.0106 ± 0.0001
Sicily west	0.2761 ± 0.0021	0.0122 ± 0.0041	0.0292 ± 0.0054		0.3209 ± 0.0049	0.0319 ± 0.0006	0.0318 ± 0.0012
Majorca	0.4162 ± 0.0087	0.0164 ± 0.0044			0.2495 ± 0.0080	0.0452 ± 0.0019	

	Haplotypes without the *Alu* insertion				
	5 (85)	6 (90)	8 (100)	10 (110)	Global GD
Sardinia C	0.0128 ± 0.0042	0.2816 ± 0.0059		0.0125 ± 0.0038	0.7751 ± 0.0254
Sardinia WC		0.2214 ± 0.0053			0.8154 ± 0.0227
Corsica C	0.0256 ± 0.0076	0.3068 ± 0.0045			0.7881 ± 0.0238
Corsica WC		0.1971 ± 0.0030			0.7387 ± 0.0338
Sicily east	0.0352 ± 0.0059	0.3330 ± 0.0094	0.0211 ± 0.0012	0.0253 ± 0.0074	0.8193 ± 0.0196
Sicily west	0.0111 ± 0.0021	0.2431 ± 0.0041		0.0408 ± 0.0048	0.8211 ± 0.0194
Majorca		0.2564 ± 0.0044			0.7501 ± 0.0369

STR alleles are expressed in number of repeats; parentheses show the size in base pairs. Haplotypes with frequencies lower than 1% are excluded

Table 4 F13B *Alu*/short tandem repeat (STR) haplotype frequencies and global gene diversities

	Haplotypes with the *Alu* insertion			
	7 (176)	8 (180)	9 (184)	10 (188)
Sardinia C		0.2503 ± 0.0143	0.0529 ± 0.0105	0.0846 ± 0.0129
Sardinia WC		0.1329 ± 0.0083	0.1296 ± 0.0114	0.0985 ± 0.0128
Corsica C	0.0205 ± 0.0033	0.3056 ± 0.0063	0.0585 ± 0.0095	0.0592 ± 0.0086
Corsica WC		0.3925 ± 0.0125	0.1094 ± 0.0092	0.0813 ± 0.0087
Sicily east	0.0341 ± 0.0001	0.2800 ± 0.0148	0.0623 ± 0.0120	0.0895 ± 0.0114
Sicily west	0.0109 ± 0.0001	0.2028 ± 0.0185	0.0353 ± 0.0101	0.1310 ± 0.0170
Majorca		0.2729 ± 0.0061	0.1266 ± 0.0151	0.0603 ± 0.0144

	Haplotypes without the *Alu* insertion				
	6 (172)	8 (180)	9 (184)	10 (188)	Global GD
Sardinia C	0.1101 ± 0.0036	0.0552 ± 0.0143	0.1554 ± 0.0105	0.2904 ± 0.0129	0.8279 ± 0.0237
Sardinia WC	0.0832 ± 0.0012	0.0477 ± 0.0083	0.2592 ± 0.0114	0.2487 ± 0.0128	0.8294 ± 0.0204
Corsica C	0.0821 ± 0.0050		0.1224 ± 0.0095	0.3450 ± 0.0086	0.7763 ± 0.0273
Corsica WC	0.0832 ± 0.0011	0.0450 ± 0.0125	0.0468 ± 0.0092	0.2416 ± 0.0087	0.7726 ± 0.0288
Sicily east	0.1136 ± 0.0002	0.0722 ± 0.0148	0.0967 ± 0.0117	0.2514 ± 0.0114	0.8239 ± 0.0217
Sicily west	0.0865 ± 0.0021	0.1776 ± 0.0185	0.0951 ± 0.0101	0.2603 ± 0.0170	0.8239 ± 0.0193
Majorca	0.1157 ± 0.0037		0.1180 ± 0.0151	0.2907 ± 0.0144	0.8236 ±0.0206

STR alleles are expressed in number of repeats; parentheses show the size in base pairs

Heterogeneity within west-Mediterranean islands has been examined in a wider context (Fig. 3a) to determine its true significance. F_{ST}-related genetic distances among our samples and a set of related populations for 8 *Alu* polymorphisms ranged from the lowest value of 0.0018 between two Spanish samples (northeast Spain and Navarre) to the highest 0.1379 (between west-coast Sardinia and west Sicily). Average genetic distances inside west-

Table 5 DM *Alu*/short tandem repeat (STR) haplotype frequencies and global gene diversities

	Haplotypes with the *Alu* insertion							
	5 (77)	10 (92)	11 (95)	12 (98)	13 (101)	14 (104)	15 (107)	20 (122)
Sardinia C	0.3440 ± 0.0228	0.0357 ± 0.0001				0.1071 ± 0.0001	0.0276 ± 0.0149	0.0357 ± 0.0001
Sardinia WC	0.2352 ± 0.0252							
Corsica C	0.5292 ± 0.0064	0.0217 ± 0.0001		0.0110 ± 0.0012		0.0679 ± 0.0068		
Corsica WC	0.3843 ± 0.0081	0.0222 ± 0.0001	0.0111 ± 0.0001	0.0135 ± 0.0056	0.0249 ± 0.0053	0.0333 ± 0.0001		
Sicily east	0.4280 ± 0.0108		0.0272 ± 0.0049	0.0466 ± 0.0103	0.0584 ± 0.0091	0.0504 ± 0.0032	0.0128 ± 0.0001	0.0183 ± 0.0110
Sicily west	0.4123 ± 0.0113	0.0236 ± 0.0041	0.0236 ± 0.0043			0.0262 ± 0.0114	0.0152 ± 0.0053	0.0109 ± 0.0001
Majorca	0.2815 ± 0.0179		0.0252 ± 0.0121	0.0492 ± 0.0039	0.0483 ± 0.0152	0.0290 ± 0.0070		0.0161 ± 0.0043

	21 (125)	22 (128)	24 (134)	25 (137)	26 (140)	28 (146)	Global GD
Sardinia C	0.0357 ± 0.0001						0.8466 ± 0.0534
Sardinia WC		0.0500 ± 0.0001					0.8789 ± 0.0432
Corsica C	0.0326 ± 0.0002					0.0109 ± 0.0001	0.7504 ± 0.0418
Corsica WC	0.0111 ± 0.0001	0.0111 ± 0.0001				0.0111 ± 0.0001	0.8005 ± 0.0333
Sicily east		0.0893 ± 0.0023		0.0253 ± 0.0021	0.0128 ± 0.0001		0.7862 ± 0.0428
Sicily west	0.0217 ± 0.0001			0.0109 ± 0.0001			0.8051 ± 0.0350
Majorca	0.0220 ± 0.0148	0.0103 ± 0.0111	0.0161 ± 0.0058				0.8719 ± 0.0194

	Haplotypes without the *Alu* insertion							
	5 (77)	11 (95)	12 (98)	13 (101)	14 (104)	15 (107)	20 (122)	21 (125)
Sardinia C	0.1203 ± 0.0228	0.0357 ± 0.0001	0.1329 ± 0.0190	0.0682 ± 0.0103				
Sardinia WC	0.0648 ± 0.0252	0.0500 ± 0.0001	0.2409 ± 0.0193	0.2000 ± 0.0001	0.0500 ± 0.0001	0.0943 ± 0.0159		
Corsica C		0.0652 ± 0.0008	0.0433 ± 0.0012	0.0972 ± 0.0024	0.0734 ± 0.0068	0.0109 ± 0.0001		
Corsica WC	0.0713 ± 0.0081	0.0444 ± 0.0001	0.0754 ± 0.0056	0.1974 ± 0.0053	0.0333 ± 0.0001		0.0111 ± 0.0001	
Sicily east		0.0497 ± 0.0049	0.0816 ± 0.0103	0.0698 ± 0.0091				
Sicily west	0.0116 ± 0.0113	0.1068 ± 0.0043	0.0958 ± 0.0047	0.1181 ± 0.0040	0.0173 ± 0.0114		0.0109 ± 0.0001	0.0109 ± 0.0001
Majorca	0.0958 ± 0.0179	0.0502 ± 0.0121	0.1017 ± 0.0039	0.1498 ± 0.0152	0.0181 ± 0.0070			0.0157 ± 0.0148

STR alleles are expressed in number of repeats; parentheses show the size in base pairs. Haplotypes with frequencies lower than 1% are excluded

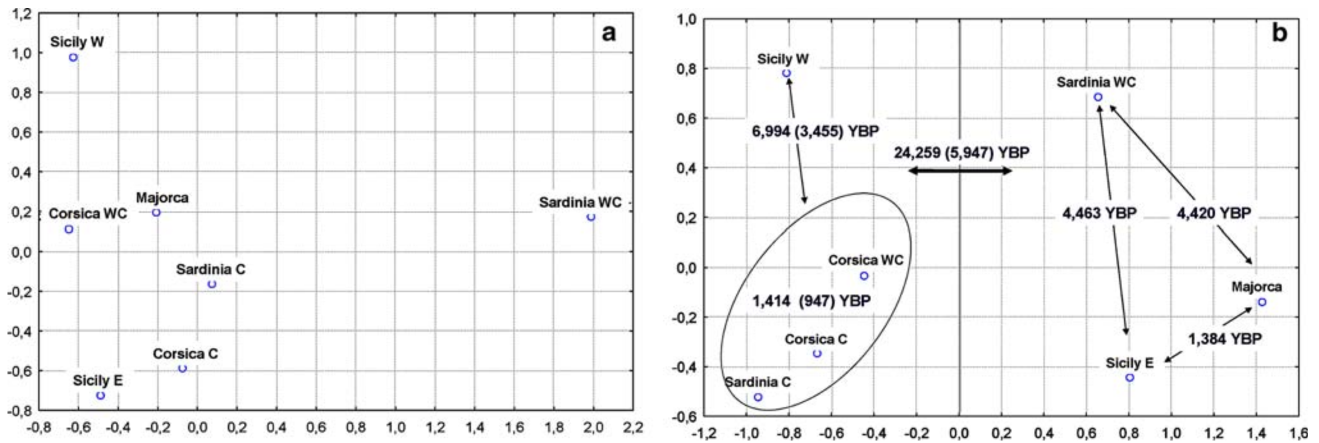

Fig. 2 **a** Plot of multidimensional scaling (MDS) (stress = 0.008) applied to the F_{ST} genetic distance matrix based on 18 *Alu* markers. **b** Plot of MDS (stress < 0.001) applied to the $(\delta\mu)^2$ genetic distance matrix based on three short tandem repeats (STRs). Years before present (YBP) estimated through the distance values are indicated for the main groups; for averaged YBP, standard deviations (SDs) are indicated in *parentheses*

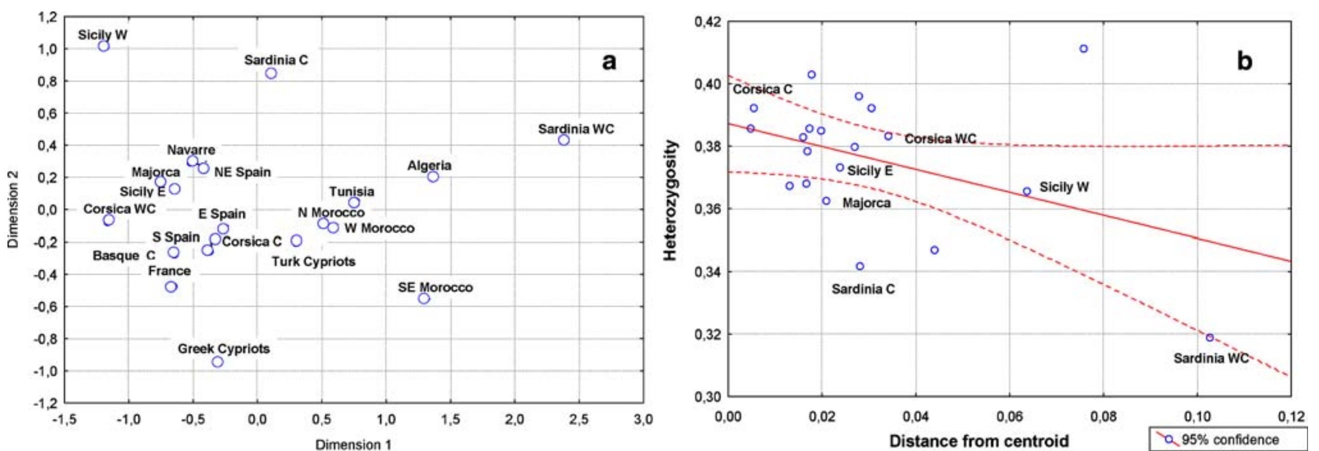

Fig. 3 **a** Plot of multidimensional scaling (MDS) (stress = 0.091) applied to the F_{ST} genetic distance matrix based on eight *Alu* markers. **b** Position of west-Mediterranean islands in the heterozygosity vs. distance from the centroid plot based on *Alu* polymorphisms

Mediterranean islands (average distance dm = 0.044) were considerably higher than the average distances among southwestern Europeans (dm = 0.010) and North Africans (dm = 0.013). On average, west-Mediterranean islands showed the highest between-group distance with North Africans (0.045). The fraction of genetic variance resulting from differences among these two groups measured through the across-loci average F_{CT} value was 1.79% ($p < 0.001$).

When population relationships were depicted through an MDS plot (Fig. 3a), west-coast Sardinia, central Sardinia and west Sicily occupied a peripheral position in the upper part of the graphic, whereas the other samples were closely related to the Spanish and French samples. The Sardinian differentiation may be explained by the fact that, in comparison with the whole population correlation between distance from centroid and heterozygosity (Fig. 3b), they showed less heterozygosity than that expected under the Harpending and Ward (1982) model, suggesting either a greater influence of genetic isolation or smaller effective population size.

Discussion

For west-Mediterranean islands, the autosomal *Alu* and STR data reported here are the first to be described and jointly discussed in order to shed light on some of the most controversial issues of west-Mediterranean population relationships, namely, the internal degree of heterogeneity within islands, the particular affinities and/or differences among islands, the amount of external gene flow received and finally, the divergence times among these regions.

Concerning *Alu* markers, the seven west-Mediterranean samples show noticeable levels of genetic diversity, with

the only exceptions being east Sicily and west-coast Sardinia, wich have the lowest average heterozygosities. Genetic differentiation inside and among islands is extremely high as can be deduced from both the results of pairwise population comparisons (17 out of 21 cross-loci population comparisons are statistically significant) and global F_{ST} values. A general trend to low gene diversity in Sardinia (Fig. 3b) joined with discrepant patterns of *Alu* allele frequencies among samples could be consistent with such differentiation. The global degree of differentiation among islands (2.2%, $p < 0.001$) is even slightly higher than that reported in Europeans (1.9%) or North Africans (1.5%, Comas et al. 2000; 2.3%, Gonzalez-Perez et al. 2003) for a comparable set of *Alu* markers and samples.

West-coast Sardinia and west Sicily are clearly differentiated from all samples (Figs. 2a, 2b, 3a). The action of genetic drift in relatively small population groups could have contributed to their differentiation. Although we cannot ignore that historical, linguistic and some genetic evidence (Piazza 1988) in Sardinia point to differences in population settlements among central and coastal areas due to the confinement of the original population, the Nuragici, into the internal regions as a result of the Carthaginian and Roman invasions. In the case of Sicily, historical records also indicate an important retreat of the original Sicanian population due to the arrival of the continental Italian Sicels. Our results concur with evidence based on Y-chromosome haplotypes (Francalacci et al. 2003) and mtDNA (Morelli et al. 2000) that point out Sardinia and Sicily as the most differentiated populations in the west-Mediterranean basin.

The relative heterogeneity among the remaining insular samples revealed by the plot based on 18 *Alu* markers (Fig. 2a) is less evident when other Mediterranean groups are added to the MDS analysis. The proximity of Corsica, Majorca and east Sicily to continental samples indicated by the set of eight *Alu* markers (Fig. 3a) has also been suggested by data from mtDNA (Falchi et al. 2006; Picornell et al. 2005).

Genetic differentiation in west-Mediterranean islands is not evident by STR variation. Although all samples show notable levels of within-population diversity, neither significant population differences nor remarkable levels of genetic variance have been detected in any of the three analysed STRs. The discrepancies observed between results indicated by *Alu* markers and STR variation may derive from the different nature of these two polymorphisms. The former are unique events far from the effect of random fluctuations caused by mutation and probably reflect the ancestral origin of populations. A population split from this ancestral group with enough time to accumulate STR variation due to the high microsatellite mutation rates, together with the homogenising effect of gene flow, could explain the observed discrepancies in genetic heterogeneity and F_{ST} values among these two genetic markers.

Gene flow among west-Mediterranean islands and beyond seems to have been outstanding. STR variation on the three loci coincide in showing high heterozygosity values in all samples; in most cases, STR variation parameters are higher than those reported for mainland Europeans (Tishkoff et al. 1996, 1998; Esteban et al. 2004). Insularity has not acted as a strong barrier to gene flow, at least among west-Mediterranean islands and mainland southern Europe, according to the merged historical background of these samples. However, historical records also point out North African influences. We have not detected any remarkable affinity among west-Mediterranean islands and North Africans. But this fact does not exclude some particular examples of African gene flow. Traces of African contributions to the gene pool of some islands can be deduced from the frequency of several CD4 *Alu*/STR haplotypes. The relatively high contribution of African-characteristic haplotypes in Sicily (8.16% in the east sample and 5.16% in the west sample) in comparison with the other islands (less than 2.5%) agrees with the strategic geographic position and the historical background of this island. Majorca, however, which was under Islamic rule for more than three centuries, does not exhibit any trace of African haplotypes. This fact agrees with other genetic data (Picornell et al. 2005) reinforcing the historical evidence that documented an important repopulation of the island by Spaniards after the Catalano–Aragonese conquest.

We conclude with some data of divergence times among samples, even though these estimations represent maximum values, because they are based on the assumption that the measured STR variation has developed locally, and we know that gene-flow processes in the west Mediterranean could have added some bias to time calculations. Time estimates (Fig. 2b) separate our samples by a time range of around 24,259 ± 6,211 YBP in two groups: central Sardinia, central Corsica, west-coast Corsica and west Sicily vs. west-coast Sardinia, east Sicily and Majorca. An average date of 5,973 ± 2,815 YBP separates west Sicily and west-coast Corsica from the remaining populations inside their respective groups. These dates are compatible with the population heterogeneity revealed by *Alu* data, suggesting that some differences among our samples could be traced back to the first settlement of the islands, likely reflecting genetic drift and/or genetic isolation processes. On the other hand, the high within-population diversities and the remarkable STR and *Alu*/STR homogeneity among islands suggest that, at least since Neolithic times, gene flow has been active in the west-Mediterranean basin. Genetic drift in west-coast Sardinia and gene flow in west Sicily have probably stressed their general genetic differentiation.

Acknowledgments We thank all of the anonymous islanders for their participation in the study, and all those who contributed to the sampling, for their valuable collaboration. This work was supported by grant CGL2005–03391 from the Spanish Ministry and grant 2005SGR00252 from the Generalitat de Catalunya. The work of EGP was financed by grant 2001FI00177 from the Generalitat de Catalunya. We give special thanks to A.D. Hadley Loera for careful revision of the manuscript.

References

Belkhir K, Borsa P, Chikhi L, Rafauste N, Bonhomme F (1996–2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France)

Black WC, Krafsur E (1985) A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. Theor Appl Genet 70:491–496

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K et al (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein family member. Cell 68:799–808

Calafell F, Bertranpetit J, Rendine S, Cappello N, Mercier P, Amoros JP, Piazza A (1996) Population history of Corsica: a linguistic and genetic analysis. Ann Hum Biol 23:237–251

Calo CM, Garofano L, Mameli A, Pizzamiglio M, Vona G (2003) Genetic analysis of a Sicilian population using 15 short tandem repeats. Hum Biol 75:163–178

Calo CM, Piras IS, Moral P, Falchi A, Ghiani ME, Varesi L, Vona G (2005) Analisi Molecolare delle popolazioni del Mediterraneo traverso 11 inserzioni Alu. Antropo 9:1–12. http://www.didac.ehu/antropo

Cappello N, Rendine S, Griffo R, Mameli GE, Succa V, Vona G, Piazza A (1996) Genetic analysis of Sardinia: I. Data on 12 polymorphisms in 21 linguistic domains. Ann Hum Genet 60:125–141

Chakraborty R, Kimmel M, Stivers DN, Davidson LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 95:1041–1046

Comas D, Calafell F, Berchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A (2000) Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Strait. Hum Genet 107:312–319. doi:10.1007/s004390000370

Day J (1987). Gli uomini e il territorio: i grandi orientimenti del popolamento sardo dall'XI al XVIII secolo. Storia dei sardi e della Sardegna. Vol. II Milano Jaka Book, pp 13–47

Edwards MC, Gibbs RA (1992) A human dimorphism resulting from loss of an Alu. Genomics 14:590–597

Esteban E, González-Pérez E, Harich N, Lopez-Alomar A, Via M, Luna F, Moral P (2004) Genetic relationships among Berbers and South Spaniards based on CD4 microsatellite/Alu haplotypes. Ann Hum Biol 31:202–212. doi:10.1080/03014460310001652275

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

Falchi A, Giovannoni L, Calo CM, Piral IS, Moral P, Paoli G, Vona G, Varesi L (2006) Genetic history of some western Mediterranean human isolates through mtDNA HVR1 polymorphisms. J Hum Genet 51:9–14. doi:10.1007/s10038-005-0324-y

Falchi A, Giovannoni L, Vacca L, Latini V, Vona G, Varesi L (2005) Beta-globin gene cluster haplotypes associate with beta-thalassemia on Corsica island. Am J Hematol 78:27–32. doi:10.1002/ajh.20199

Felsenstein J (1989) PHYLIP-phylogeny inference package. Cladistics 5:164–166

Flores C, Maca-Meyer N, Gonzalez AM, Cabrera VM (2000) Northwest African distribution of the *CD4/Alu* microsatellite haplotypes. Ann Hum Genet 64:321–327

Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, Madeddu R, Paoli G, Tofanelli S, Calo CM, Ghiani ME, Varesi L, Memmi M, Vona G, Lin AA, Oefner P, Cavalli-Sforza LL (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. Am J Phys Anthrop 121:270–279. doi:10.1002/ajpa.10265

Garcia-Obregon S, Alfonso-Sanchez MA, Perez-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. doi:10.1002/ajhb.20487

Garcia-Obregon S, Alfonso-Sanchez MA, Perez-Miranda AM, De Pancorbo MM, Peña JA (2007) Polymorphic *Alu* insertions and the genetic structure of Iberian Basques. J Hum Genet 52:317–327. doi:10.1007/s10038-007-0114-9

Gatti AM (1995) Tra demografia e genetica: popolazione e matrimonio in Sardegna. Antropologia Contemporanea 18:67–80

Ghiani ME, Calo CM, Autuori L, Mameli GE, Succa V, Vacca L, Cerutti N, Rabino Massa E, Vona G (2002) New data on the genetic structure of the population of Sicily: analysis of the Alia population (Palermo, Italy). Am J Hum Biol 14:289–299. doi:10.1002/ajhb.10036

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

González-Pérez E, Via M, Esteban E, Lopez-Alomar A, Mazieres S, Harich N, Kandil M, Dugoujon JM, Moral P (2003) Alu insertions in the Iberian Peninsula and NW Africa: genetic boundaries or melting pot? Coll Antropol 27:491–500

Guo S, Thomson E (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372

Harpending H, Ward RH (1982) Chemical systematics and human population. In: Nitecki M (ed) Biochemical aspects of evolutionary biology. University of Chicago Press, Chicago, pp 213–252

Memmi M, Moral P, Calo CM, Autuori L, Mameli GE, Succa V, Varesi L, Vona G (1998) Genetic structure of southwestern Corsica (France). Am J Hum Biol 10:567–577

Morelli L, Grosso MG, Vona G, Varesi L, Torroni A, Francalacci P (2000) Frequency distribution of mitochodrial haplogroups in Corsica and Sardinia. Hum Biol 72:585–595

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nishimura DY, Murray JC (1992) A tetranucleotide repeat for the F13B locus. Nucleic Acids Res 20:1167

Piazza A, Cappello N, Olivetti E, Rendine S (1988) A genetic history of Italy. Ann Hum Genet 52:203–213

Picornell A, Gomez-Barbeito L, Tomas C, Castro JA, Ramon MM (2005) Mitochondrial DNA HVRI variation in Balearic populations. Am J Phys Anthrop 128:119–130. doi:10.1002/ajpa.10423

Rajeevan H, Cheung KH, Gadagkar R, Stein S, Soundararajan U, Kidd JR, Pakstis AJ, Miller PL, Kidd KK (2005) ALFRED: an allele frequency database for microevolutionary studies. Evolut Bioinf 1:1–10

Ramakrishnan U, Mountain J (2004) Precision and accuracy of divergency time estimates from STR and SNPSTR variation. Mol Biol Evol 21:1960–1971. doi:10.1093/molbev/msh212

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

Rickards O, Martinez-Labarga C, Scano G, De Stefano GF, Biondi G, Pacaci M, Walter H (1998) Genetic history of the population of Sicily. Hum Biol 70:699–714

Romano V, Cali F, Ragalmuto A, D'Anna RP, Flugy A, De Leo G, Giambalvo O, Lisa A, Fiorani O, Di Gaetano C, Salerno A, Tamouza R, Charron D, Zei G (2003) Autosomal microsatellite and mtDNA genetic analysis in sicily (Italy) Ann Hum Genet 67:42–53. doi:10.1046/j.1469-1809.2003.00007.x

Simi P (1997) L'apport de la documentation de la Franciscorsa à la connaissance de la démographie de la Corse. Bull Soc Hist Nat Corse 679:25–43

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

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction. Am J Hum Genet 73:1162–1169. doi:0002-9297/2003/7305-0018$15.00

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989. doi:0002-9297/2001/6804-0020$02.00

Stoneking M, Fontius JJ, Clifford S, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071.doi: 10.1101/gr.7.11.1061

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M, Pääbo S, Watson E, Risch N, Jenkins T, Kidd KK (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380–1387. doi: 10.1126/science.271.5254.1380

Tishkoff SA, Goldmans A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK (1998) A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402. doi:0002-9297/98/6206-0016$02.00

Tofanelli S, Taglioli L, Varesi L, Paoli G (2001) STR-based genetic profile of the population of Corsica island (France). Forensic Sci Int 121:3–38. doi:10.1016/S0379-0738(01)00516-3

Vona G, Moral P, Memmi M, Ghiani ME, Varesi L (2003) Genetic structure and affinities of the Corsican population (France): classical genetic markers analysis. Am J Hum Biol 15:151–163. doi:10.1002/ajhb.10133

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. doi:0002-9297/2001/6803-0018$02.00

Supplementary material

Table of CD4, F13B and DM STR variation

	Sard C	**Sard WC**	**Cor C**	**Cor WC**	**Sic E**	**Sic W**	**Majorca**
Alleles*				*CD4 STR variation*			
2N	90	94	94	96	94	96	100
5	0.3889	0.3617	0.3723	0.4375	0.3298	0.2917	0.4300
6	0.3444	0.2447	0.2979	0.1979	0.3830	0.2500	0.2600
7	0.0000	0.0000	0.0213	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0213	0.0313	0.0000
9	0.0000	0.0106	0.0000	0.0104	0.0000	0.0000	0.0000
10	0.2444	0.2979	0.2553	0.3021	0.2128	0.3646	0.2600
11	0.0000	0.0532	0.0532	0.0417	0.0426	0.0313	0.0500
12	0.0222	0.0319	0.0000	0.0104	0.0106	0.0313	0.0000
				F13B STR variation			
2N	72	76	94	96	88	92	94
6	0.1111	0.0921	0.0851	0.0833	0.1136	0.0870	0.1170
7	0.0000	0.0000	0.0213	0.0000	0.0341	0.0109	0.0000
8	0.3056	0.2105	0.3085	0.4375	0.3523	0.3804	0.2766
9	0.2083	0.3684	0.1809	0.1563	0.1591	0.1304	0.2447
10	0.3750	0.3289	0.4043	0.3229	0.3409	0.3913	0.3511
11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106
				DM STR variation			
2N	94	94	96	96	94	96	104
5	0.4255	0.3723	0.5104	0.4583	0.4149	0.4479	0.3684
8	0.0000	0.0000	0.0104	0.0104	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0280	0.0000
10	0.0532	0.0213	0.0208	0.0208	0.0000	0.0417	0.0000
11	0.0957	0.0957	0.0625	0.0625	0.0851	0.1354	0.0789
12	0.2128	0.2128	0.0521	0.0938	0.1277	0.0833	0.1491
13	0.0851	0.1064	0.1250	0.2188	0.1383	0.1146	0.2018
14	0.0426	0.0106	0.1354	0.0625	0.0532	0.0417	0.0526
15	0.0106	0.0213	0.0104	0.0000	0.0106	0.028	0.0088
16	0.0000	0.0106	0.0000	0.0000	0.0106	0.0000	0.0175
17	0.0106	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
18	0.0000	0.0213	0.0000	0.0000	0.0106	0.0000	0.0000
> 18	0.0637	0.1275	0.0729	0.0728	0.1489	0.0942	0.1128

* Alleles are expressed in number of repeats; 2N = number of chromosomes.

12 *Genetic Change in the Polynesian Population of Easter Island: Evidence from* Alu *Insertion Polymorphisms*

Emili González-Pérez, Esther Esteban, Marc Via, Clara García-Moro,
Miquel Hernández & Pedro Moral.

Annals of Human Genetics 2006; 70: 829-840

12.1 *Canvi genètic a la població polinèsica d'Illa de Pasqua: evidències aportades per polimorfismes d'inserció* Alu

El darrer dels treballs experimentals recollits en els resultats d'aquesta tesi doctoral utilitza l'aproximació ja comentada anteriorment a partir de l'anàlisi genètica d'una bateria de polimorfismes genètics *Alu*, per ser aplicada a l'estudi antropogenètic i evolutiu d'una població humana molt particular. Es tracta d'un estudi novedós sobre les característiques biològiques de la població autòctona de l'Illa de Pasqua, sobre les seves vicissituds històriques recents i sobre les possibles inferències per intentar clarificar l'origen dels pobladors polinèsics i el model de poblament més escaient.

L'origen dels pobladors i habitants de les nombroses illes de la Polinèsia i l'Oceà Pacífic ha estat durant molt de temps, i encara actualment, un dels temes més controvertits i difícils d'afrontar per l'antropologia i la genètica de poblacions humanes. Amb la intenció de reconduir aquesta problemàtica, en el decurs del projecte d'aquesta tesi es va considerar interessant aplicar l'estudi de nous marcadors autosòmics com les insercions *Alu*–mai estudiats abans per a la població de l'Illa de Pasqua– per tal d'obtenir pistes sobre la història recent i l'origen dels pobladors de la Polinèsia més remota.

Així, van ser analitzades un total de 18 polimorfismes d'inserció *Alu* en una completa mostra humana de l'Illa de Pasqua, constituïda per un total de 88 individus nadius. Aquesta població global va ser subdividida en dos grups, en funció de les genealogies disponibles als registres religiosos de l'illa des de finals del segle XIX: una subpoblació estava formada per individus 100 % descendents de llinatges rapanui (els

pobladors nadius originals d'Illa de Pasqua) i una altra subpoblació configurada per individus que en les darreres quatre generacions haurien tingut algun ancestre no rapanui (europeu o amerindi, principalment). Aquesta particular caracterització inicial de la població analitzada ens va permetre disposar d'una mostra equivalent als illencs habitants de l'illa a finals del segle XIX (individus rapanui, sense pràcticament mestissatge extern teòric) i una mostra característica de la barreja genètica que ha patit la població original, especialment des de mitjans del segle XX.

L'anàlisi genètica ha permès confirmar, a una escala biològica, aquesta diferenciació entre els dos grups d'habitants de l'illa, permetent-nos utilitzar els resultats per tal d'inferir processos demogràfics i històrics sobre l'origen i l'evolució d'aquesta població insular aïllada al bell mig del Pacífic.

Les dades demostren una participació equivalent de gens externs amerindis i europeus en la població autòctona de l'illa de finals del segle XIX, i amb una creixent i accelerada contribució del *pool* genètic europeu durant el segle XX, especialment marcada a partir de la dècada dels seixanta. Cal destacar, a més, que aquestes dades mostren l'evident petjada genètica ja deixada en la població nativa original (finals del segle XIX) pels colonitzadors europeus i per l'efecte de l'intens comerç d'esclaus polinèsics durant aquell segle, una petjada i un contacte que haurien tingut un paper decisiu en el declivi de la població i que la va abocar a gairebé una extinció en tota regla.

La comparació de la variació genètica d'elements *Alu* obtinguda en aquesta població amb les dades disponibles per altres poblacions de la regió donen un suport clarament preferent als models de poblament del Pacífic anomenats del *Voyaging Corridor*, que postulen que els humans prepolinesis tindrien el seu origen principal en el sud-est asiàtic i en les poblacions humanes de la Wallacea, més que no pas en poblacions de l'Àsia oriental estricta (Taiwan i Filipines) que havien estat els candidats tradicionalment postulats per molts antropòlegs.

Aquest estudi ha permès demostrar la importància que el procés sistemàtic de mostreig i la caracterització exhaustiva de la població durant el mateix, pot arribar a tenir en la correcta interpretació de les relacions que mostren els resultats de l'anàlisi genètica en detall. Així, ha estat possible desentrellar algunes de les particularitats que han afectat a la història recent d'una població humana tan particular com l'Illa de Pasqua i com es va desenvolupar el seu procés de poblament més probable.

12.2 Informe del director sobre la participació del doctorand

El Dr. **Pedro Moral Castrillo**, Professor Titular del Departament de Biologia Animal de la Universitat de Barcelona i director de la tesi doctoral **"Variació genètica i evolució d'elements *Alu* recents en poblacions humanes. Inferències biodemo-gràfiques i filogeogràfiques"** presentada pel doctorand **Emili González Pérez**, desitja fer constar que la participació del doctorand en l'elaboració de l'article **"Genetic Change in the Polynesian Population of Easter Island: Evidence from *Alu* Insertion Polymorphisms"** publicada per la revista *Annals of Human Genetics*, ha consistit en les següents tasques principals:

- Processament de mostres: extracció de DNA de les mostres poblacionals utilitzades en el treball (100 %)
- Disseny del treball i selecció de marcadors analitzats, conjuntament amb el Dr. Pedro Moral
- Genotipatge de marcadors d'inserció *Alu* i seqüenciacions (100 %)
- Elaboració de les bases de dades de resultats (100 %)
- Anàlisi estadística dels resultats (100 %)
- Redacció de l'article, conjuntament amb el Dr. Moral

Complementàriament, cal assenyalar que cap dels coautors d'aquest article ha utilitzat, ni implícitament ni explícitament, els resultats d'aquest treball per a l'elaboració d'una altra tesi doctoral. En conseqüència, aquest treball forma part, en exclusiva, del treball de recerca en el que s'emmarca la tesi del doctorand Emili González Pérez.

Signat: Dr. Pedro Moral Castrillo
Barcelona, 29 de setembre de 2009

doi: 10.1111/j.1469-1809.2006.00293.x

Genetic Change in the Polynesian Population of Easter Island: Evidence from *Alu* Insertion Polymorphisms

E. González-Pérez, E. Esteban, M. Via, C. García-Moro, M. Hernández and P. Moral*

Unitat d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Barcelona, Spain

Summary

The origin of Pacific islanders is still an open issue in human population genetics. To address this topic we analyzed a set of 18 *Alu* insertion polymorphisms in a total of 176 chromosomes from native Easter Island inhabitants (*Rapanui*). Available genealogical records allowed us to subdivide the total island sample into two groups, representative of the native population living in the island around 1900, and another formed by individuals with some ancestors of non-*Rapanui* origin. Significant genetic differentiation was found between these groups, allowing us to make some biodemographic and historical inferences about the origin and evolution of this geographically isolated island population. Our data are consistent with equivalent and recent contributions from Amerindian and European migrants to the 1900s *Rapanui* population, with an accelerated increase in the European gene flow during the 20th century, especially since the 1960s. Comparative analysis of our results with other available *Alu* variation data on neighbouring populations supports the "Voyaging Corridor" model of Polynesian human settlement, which indicates that pre-Polynesians are mainly derived from Southeast Asian and Wallacean populations rather than from Taiwan or the Philippines. This study underlines the importance of sampling and taking into account historical information in genetic studies to unravel the recent evolution of human populations.

Keywords: *Alu* insertions, Easter Island, Polynesian origins, Pacific peopling, human populations

Introduction

Geographically, the Polynesian Easter Island (as named by European explorers in 1722, *Isla de Pascua* in Spanish, *Rapanui* in Polynesian, and '*Te Pito o TeHenua*' or "the navel of the world" by its inhabitants) is the world's most isolated inhabited place. This volcanic triangular island lies in the Austral group in Eastern Polynesia, 2,300 miles west of South America, 2,500 miles southeast of Tahiti, 4,300 miles south of Hawaii and 3,700 miles north of Antarctica. Due to its geographic isolation and particular culture (the "moai" huge stone carved figures) it has attracted the interest of anthropologists for a long time. It is generally accepted that this island has however been colonized for only 1,500 years, in one of the most recent processes of modern human expansion.

Archaeological and linguistic data place the colonization of Easter Island around 400-500 AD, with settlers likely coming from the Marquesas Islands or Mangavera. The subsequent growth of the small initial population gave rise to a particular and complex culture centred on the cult of their ancestors (McCall, 1996). The population growth, along with the limited environmental resources and an unstable ecological equilibrium, unleashed a crisis in the 16th century, leading to an important reduction in the size of the population and a change of the dominant culture (Moai abandonment). The arrival of the Dutchman J. Roggeveen in 1722 started a period of contact with Europeans that, finally, led to the decline of the native population. It is well documented that the *Rapanui* population suffered an extreme reduction from an initial size of around 2,000 to only 110 in 1877, as a result of foreign exploitation and

*Corresponding author: Prof. Pedro Moral, Unitat d'Antropologia, Facultat de Biologia. Avda. Diagonal, 645. 08028-Barcelona, Spain. Phone: +34 934 021 461; Fax: +34 934 035 740; E-mail: pmoral@ub.edu

© 2006 The Authors
Journal compilation © 2006 University College London

exposure to introduced epidemics (Flenley, 1993). Although no genetic evidence of this bottleneck has been found (Martinson *et al.* 1993), historically this reduced population group should be regarded as the origin of the current *Rapanui*, together with likely later foreign influences. Chilean sovereignty in 1888, the sheep and cattle-raising concessions to foreigners or the arrival of German marines and English prisoners in 1915, help to illustrate these potential influences. Demographic data indicate that until 1965 the *Rapanui* population maintained a strong endogamy, and afterwards the island opened significantly, increasing its size to 2764 islanders by 1992, mainly due to Chilean immigration (Hernández *et al.* 2000).

The reconstruction of Polynesian ancestry has been approached by archaeologists, linguists and genetic anthropologists leading to different conflicting hypotheses that make this issue a highly disputed topic in the elucidation of the evolution of present-day human populations. The generation of new data on the genetic structure of the most extreme population within Polynesia, the *Rapanui*, may be useful in answering some of the questions still raised by this controversy.

Among the different hypotheses (see, for instance, Diamond, 1988; Dyen, 1971; Heyerdahl, 1950; Oppenheimer & Richards, 2001; Terrell & Welsch, 1997) two basic models of Polynesian origins currently have major support. The 'Out of Taiwan' model (also known as the 'Express Train to Polynesia') associates the ancestors of Polynesians with a fast Neolithic expansion of Austronesian-speaking rice agriculturalists from South China/Taiwan, and is mainly based on linguistic and cultural evidence. A modified version of this model, the "Slow Boat" model (Kayser *et al.* 2000), proposes a Taiwan or East Asia origin for Polynesians with extensive admixture with Melanesians. The main alternative model, the 'Voyaging Corridor', postulates somewhere in the Southeast Asia islands (the *Wallacea* region) as the homeland of today's Polynesian ancestors, with a demographically more stable spreading, with contacts made through maritime journeys, initially at short distances and later at greater distances. Finally, the model called "Entangled Bank" suggests a primarily Melanesian origin for Polynesians (Terrell & Welsch, 1997).

A good number of previous anthropological and molecular studies focused on the topic of human settlement of the Pacific. Cranial morphology data have suggested similarities between Polynesians and SE Asians (Pietrusewsky, 1997), as well as closer affinities of Paleoamericans with SE Asians and Polynesia than with mainland Asia (Brace *et al.* 2001; González-José *et al.* 2003). Most genetic evidence is consistent with a proximal biological origin for Pacific populations in the *Wallacea* region (Serjaeantson & Hill, 1989), including globin genes (Hill *et al.* 1985), Y-chromosome markers (Kayser *et al.* 2000; Capelli *et al.* 2001; Hurles *et al.* 2002) and some mtDNA studies (Sykes *et al.* 1995; Oppenheimer & Richards, 2001). Other mtDNA data have been interpreted as supporting affinities of Polynesians with Taiwanese populations (Redd *et al.* 1995; Melton *et al.* 1995; Lum *et al.* 1998) and these researchers insist there is still the possibility of limited gene flow between Native Americans and Polynesian populations (Cann & Lum, 1996). In this context the present analysis, dealing with up-to-date and previously unused markers, aims to find new genetic evidence to test the origin of Polynesians.

Here we provide, for the first time, data on 18 *Alu* insertion polymorphisms in the Easter Island population. The use of neutral autosomal polymorphisms, such as *Alu* insertions, whose frequency variation is only dependent on drift and migration (and not mutation), and whose ancestral state is known, is a powerful tool in the study of human populations (Watkins *et al.* 2001). These markers have been determined in two genealogically well documented subsamples of the current island population, which can be associated with two historical periods of this population. The objectives of this work are: 1) to investigate the autosomal genetic composition of the Easter Island populations; 2) to assess the amount of change in the gene pool of this island population that is attributable to recent demographic events; 3) to detect genetic affinities with other geographically and/or historically related populations, in order to gain new insights into the general model of Polynesian human settlement.

Materials and Methods

A group of 88 unrelated autochthonous individuals from Easter Island was analysed. The samples were obtained in collaboration with the authorities of the Hanga

© 2006 The Authors
Journal compilation © 2006 University College London

Figure 1 Location of population samples used for comparison purposes in this study. *Alu* frequency data sources are indicated in the text. Abbreviations: RAPA = RR *Rapanui* sample, SAMO = Samoa, JAVA = Java, PHIL = Philippines, TAIW = Taiwan, TENG = Nusa Tengaras, MALA = Malasia, CHIN = China, TAMI = Tamil, INDI = India, AUST = Australia, PNGH = Papua New Guinea Highlands, PNGC = Papua New Guinea Coastal, ALAS = Alaska, MVSK = Mvskoke, MAYA = Maya, BRET = Bretons, FRAN = France, SPAI = Spain, SWIS = Swiss.

Roa Hospital and came from unrelated healthy blood donors. Informed consent was obtained from all subjects participating and the study was approved by the ethical committee of the University of Barcelona. The available genealogical records for these individuals, coming from family reconstructions from parish and civil archives in *Hanga Roa* between 1937 and 1996, document their grandparental origins and allowed subdivision into two subsamples: one formed by subjects of *Rapanui* ancestry (n = 54) and the other comprising individuals with some non-*Rapanui* ancestor (n = 34).

Eighteen human specific *Alu* insertion polymorphisms (TPA25, ACE, APO-A1, FXIIIB, PV92, D1, CD4, B65, Sb19.3, Sb19.12, A25, HS2.43, HS4.69, HS4.32, DM, Yb8NBC120, YbNBC125, and YAP) were typed using primers described previously (Arcot *et al*. 1995, 1996; Batzer & Deininger, 1991; Watkins *et al*. 2001; Edwards & Gibbs, 1992). The conditions for PCR amplification were as described in Stoneking *et al*. (1997) and Edwards & Gibbs (1992) with minor modifications.

Allele frequencies were computed by direct counting, and Hardy-Weinberg equilibrium was checked by an exact test (Guo & Thomson, 1992). Gene diversity by population and locus was calculated according to Nei's formula (Nei, 1987), and locus frequency distributions were compared by an exact test for population differentiation. Estimates were obtained using the Arlequin statistical package (Schneider *et al*. 2000).

For comparison purposes *Alu* variation data from other populations (Figure 1) were collected from the literature (Arcot *et al*. 1996; Stoneking *et al*. 1997; Melton *et al*. 1998; Novick *et al*. 1998; Battilana *et al*. 2002) in an attempt to cover the main population groups related to the origin and history of the *Rapanui*. Actually, the great dispersion of *Alu* human population data available to date led us to a mandatory compromise to maximize both the number of markers and populations. Consequently, we finally used data for 8 polymorphic *Alu* insertions from a total of 23 populations corresponding to different geographical areas: SE Asia (3 populations from the Taiwanese group: Taiwan, Philippines, and China; and 4 from the *Wallacea* region: Malaysia, Molucas, Nusa Tengara and Java); Polynesia (2 populations: Samoa and *Rapanui*); Australia and New Guinea (3 populations: Australian Aborigines, Papua New Guinea Highlands and Papua New Guinea Coast); America (2 populations from North America: Alaska and Mvskoke; 1 from Central America: Maya; and 4 from South America: Ache, Cain, Guarani, and Xavante); and Europe (4 populations: Bretons, French, Swiss, and Spaniards).

© 2006 The Authors
Journal compilation © 2006 University College London

Fst-related genetic distances were calculated between pairs of populations according to Reynolds *et al.* (1983) and represented in a neighbour joining tree (Saitou & Nei, 1987) using the PHYLIP 3.6 package (Felsenstein, 1989). The topology of the tree was assessed through 1000 bootstrap iterations. The genetic relationships between the examined populations were also depicted by principal components analysis (PCA) from the correlation matrix of the *Alu* frequencies.

Delaunay network analysis was used to identify principal boundaries or regions of sharp genetic change (Brassel & Reif, 1979). In accordance with the rules of this analytical process we defined a subset of pairs of contiguous populations connected by edges. Genetic distances between each pair of samples were allocated to each edge, and high genetic distances were joined to trace the principal genetic boundaries in the South-East Asia and Pacific region.

The proportions of genetic admixture in the *Rapanui* population were calculated by gene flow estimates obtained by the ADMIX 1.0 program (Bertorelle & Excoffier, 1998).

Results

Frequencies of the 18 human-specific *Alu* polymorphisms in the two subgroups from Easter Island (RR, all known ancestors of *Rapanui* origin, and NR, some non-*Rapanui* ancestors) are presented in Table 1. All loci were polymorphic in all populations, except for the HS2.43 locus for which the insertion allele was absent in the *Rapanui* lineage (RR). All Hardy-Weinberg tests in the two lineages studied showed no departures from equilibrium (Table 1).

Variability within the Present Rapanui Population

Gene diversity values were relatively high according to the maximum expected values for biallelic markers (Table 1). Significant differences in heterozygosity between loci were present (Kruskal-Wallis test, $p = 0.014$), but not between the two Easter Island subgroups ($p = 0.681$). However, the RR subsample showed a certain trend towards lower gene diversities (in 9 out of the 18 loci analysed, while the remaining values are equivalent) than the mixed NR subgroup. The average gene diversities by population confirmed this trend, with values of 0.287 in the RR group vs. 0.334 in the NR subsample.

When locus genotype distributions between the two Easter Island subgroups were compared significant differences were present for four of the 18 loci: DM ($p = 0.0092$), FXIIIB ($p = 0.0001$), ACE ($p = 0.0089$) and APO ($p = 0.0099$). Also a global test based on the distribution of the whole set of the markers examined yielded a significant value ($p = 0.0008$, $\chi^2 = 69.046$, d.f.$= 36$), indicating the existence of remarkable genetic heterogeneity within the present Easter Island population.

Notes on the Genetic Changes in the *Rapanui* Island Population During the Last Century

Concerning the heterogeneity detected within the current Easter Island population, genetic distance and admixture analyses can provide additional information on the genetic changes experienced in the island population during the last century and those that are likely occurring at present.

Table 2 shows a summary of the average genetic distances between the two Easter Island subsamples and population groups from SE Asia, the Taiwanese group, Australian and Melanesian group, Americans, and Europeans. The RR *Rapanui* group presented the lowest distance from SE Asians, followed by that from Americans, while the distances with other groups were almost or more than twice larger. The same distance pattern was observed for the NR *Rapanui* subsample, but in this case the magnitude of the distance values was always around a half of the RR group, in agreement with an expected loss of differentiation in the recent Easter Island population due to a migration effect.

In order to shed some light on the direction of the recent genetic changes within Easter Island, an analysis of the admixture proportions of Polynesian, American and European genes into the two subpopulations of the *Rapanui* was carried out. The respective bootstrapped contributions were $m_{Polyn} = 0.493 \pm 0.342$, $m_{Amer} = 0.211 \pm 0.334$, and $m_{Europ} = 0.296 \pm 0.115$ into the RR *Rapanui* subpopulation, and $m_{Polyn} = 0.265 \pm 0.297$, $m_{Amer} = 0.251 \pm 0.285$, and $m_{Europ} = 0.484 \pm 0.137$ into the NR *Rapanui* subgroup. Likely due to

© 2006 The Authors
Journal compilation © 2006 University College London

Table 1 *Alu* insertion frequencies in two population samples of *Rapanui* Pacific island (RR = *Rapanui* lineage and NR = Non *Rapanui* lineage) (2N sample size in number of typed chromosomes)

Alu element	Lineage	2N	*Alu* +	Heterozygosity	H-W (*p*)
TPA 25	NR	58	0.47	0.51	0.71
	RR	94	0.49	0.50	0.25
ACE	NR	58	0.69	0.44	0.21
	RR	94	0.86	0.24	1
APO A1	NR	68	0.59	0.49	0.08
	RR	108	0.41	0.48	0.40
FXIIIB	NR	62	0.65	0.47	0.44
	RR	86	0.93	0.13	0.17
PV92	NR	54	0.52	0.51	0.46
	RR	80	0.46	0.50	0.52
D1	NR	56	0.38	0.48	1
	RR	78	0.53	0.50	0.52
CD4	NR	52	0.94	0.11	1
	RR	82	0.95	0.09	1
B65	NR	62	0.45	0.50	1
	RR	84	0.49	0.50	1
YAP (N = n°♂)	NR	9	0.00	0.00	–
	RR	16	0.00	0.00	–
Sb19.3	NR	58	0.76	0.37	0.64
	RR	100	0.68	0.44	0.21
Sb19.12	NR	58	0.09	0.16	1
	RR	94	0.02	0.04	1
A25	NR	62	0.05	0.09	1
	RR	92	0.09	0.16	1
HS2.43	NR	58	0.01	0.03	0.78
	RR	86	0.00	0.00	–
HS4.69	NR	58	0.34	0.46	0.69
	RR	84	0.44	0.50	1
HS4.32	NR	62	0.39	0.48	0.45
	RR	92	0.39	0.48	0.23
DM	NR	60	0.32	0.44	0.67
	RR	94	0.15	0.25	0.57
Yb8NBC120	NR	52	0.27	0.40	0.63
	RR	74	0.15	0.26	0.57
Yb8NBC125	NR	32	0.06	0.12	1
	RR	40	0.05	0.10	1

Table 2 Average genetic distances between *Rapanui* subsamples and other population groups

	SE Asian Islands (5)[1] ($d_m = 0.051$)	Taiwanese group (3) ($d_m = 0.065$)	Australia & New Guinea (3) ($d_m = 0.096$)	Americans (3) ($d_m = 0.045$)	Europeans (4) ($d_m = 0.018$)
RR *Rapanui*	0.106	0.179	0.244	0.129	0.214
NR Easter Island	0.044	0.099	0.138	0.078	0.102
SE Asians	—	0.098	0.212	0.061	0.141
Taiwanese		—	0.328	0.116	0.170
Austr & NG			—	0.335	0.184
Americans				—	0.144

d_m: average genetic distance within each population group.
[1]in brackets the number of populations within each group.

© 2006 The Authors
Journal compilation © 2006 University College London

Table 3 AMOVA analysis: Apportionment of the *Alu* frequency variance in population groups potentially related to the *Rapanui* origin by geographic and/or historical grounds (All data with significant p-values <0.0001)

	among groups (%)	among populations within groups (%)	total between populations (%)
[1]Pop. gr. 1	11.35	5.99	17.34
[1]Pop. gr. 2	10.30	6.61	16.93

[1]Population subdivisions and included samples:
Pop gr. 1) Europe (Bretons, French, Swiss, and Spaniards), SE Asia plus Pacific (China, Taiwan, Philippines, Malaysia, Moluccas, Nusa Tengaras, Java, Samoa, *Rapanui*); Australia and New Guinea (Australia, PNG Coastal, and PNG Highland), and America (Alaska, Muskoke, Maya, Ache, Caingang, Guarani, and Xavante).
Pop gr. 2) Taiwanese group (Taiwan, Philippines, and China), SE Asian Islands plus Pacific (Malaysia, Moluccas, Nusa Tengaras, Java, Samoa, *Rapanui*), Australia and New Guinea (with the same samples than before), and America (idem).

random factors (sample sizes and stochastic variation between loci), the only statistically significant values corresponded to the European contributions. The remarkable difference in this contribution between the two Easter Island subsamples (almost twice higher in NR than RR subgroup) indicates that the introduction of European genes might have been one of the main factors in the genetic changes that occurred in the population of Easter Island during recent times.

According to these results, and in order to avoid the potential effects of migration in the last century, only the RR *Rapanui* subsample was used in the following analysis as being representative of the autochthonous population living in the island at the beginning of 1900s. In contrast, the NR *Rapanui* subgroup can be used comparatively to approach changes in the island genetic pool due to recent migration.

Genetic Origin of the Autochthonous *Rapanui* Population of the Last Century

In order to find out what *Alu* polymorphisms tell us about the origin of the *Rapanui* population our sample was compared with other geographical and/or historically related population groups from the Pacific area and Europe. According to the extant hypotheses of Polynesian ancestry, *Alu* insertion frequencies for eight loci analyzed were compiled from the literature for several samples from East Asia, Southeast Asian Islands and the *Wallacea* region, Australia and New Guinea, Polynesia and America. To check the usefulness of these markers for genetic differentiation of these population groups an AMOVA analysis was performed, in order to determine the apportionment of the allele frequency variance. This analysis was carried using different population groupings as detailed in Table 3. In the first step representatives of the major outside groups related to the ancestry of the *Rapanui*, including Europeans, were considered to get a general picture of the *Alu* variation. In the second round Europeans were excluded, and the SE Asian plus Pacific group was subdivided into a Taiwanese and other group to check for specific hypotheses on the origin of Polynesian colonization. In both cases the across loci average fraction of genetic variance resulting from differences between populations was highly significant, with values (17.31% and 16.93%, see Table 3) of the same order of magnitude as the overall values of between population genetic diversity on a worldwide scale (Cavalli-Sforza & Feldman, 2003). More interestingly, when populations were classified into groups, as specified in Table 3, the fraction of the genetic variance attributable to differences between groups was higher than that among populations within groups for most of the loci (five out of eight). In the same way the average fraction of variance among groups was higher than within groups, suggesting the usefulness of these markers for distinguishing between the established population groups. The fact that the same pattern of apportionment of the genetic variance was present in the two population subdivisions analyzed allows us to discard the possibility of a specific effect due to the inclusion of extremely differentiated populations (in this case Europeans), in the context of Pacific populations. An additional analysis of genetic diversity carried out within each population group (data not shown) indicated that Europeans were the most homogeneous group, while the *Alu*

© 2006 The Authors
Journal compilation © 2006 University College London

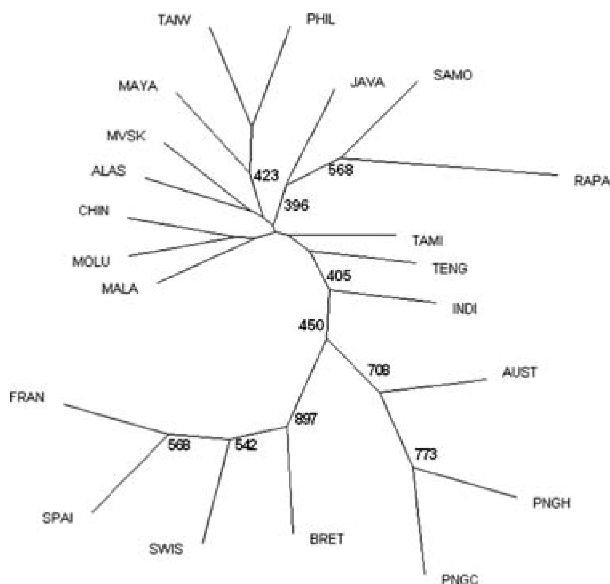

Figure 2 Neighbour-joining tree of the genetic distance matrix based on *Alu* insertion loci for *Rapanui* and other human populations. Bootstrap supports are indicated along the nodes.

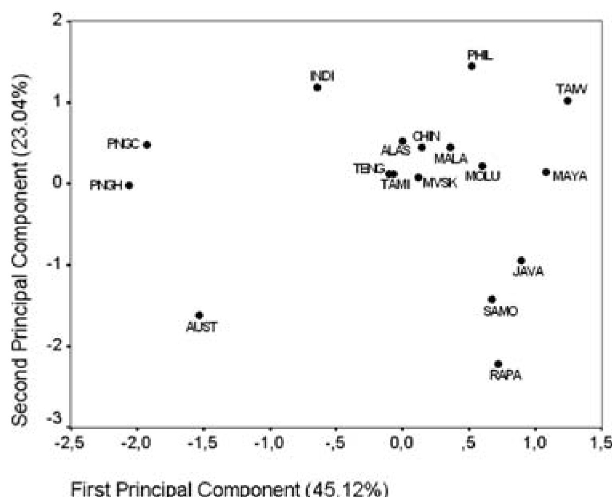

Figure 3 Plot of the two first principal components on the variation of 8 *Alu* polymorphisms in *Rapanui* and other SE Asian, Pacific and American populations.

allele distributions between the populations within the remaining groups were significantly different.

The relationships between the inhabitants of Easter Island and other human groups, assessed by Fst genetic distances, are depicted in the neighbour-joining tree in the Figure 2. The tree divides the populations into three main groups: Europe, Australia and New Guinea, and the remainder. The nodes separating the two first groups have strong bootstrap support after 1000 iterations (>70%). The third group is internally less consistent and includes the populations from SE Asia, the Pacific and America. In this cluster the SE Asian and *Wallacea* populations show an undifferentiated position, while the Taiwanese group (Taiwan and Philippines) becomes detached and potentially related with Amerindian populations (Maya). Finally, the Polynesian group (Samoa and *Rapanui*) is joined to the Java population from the *Wallacea* region.

To avoid constrictions of bifurcating models underlying any tree representation, population genetic relationships among Pacific populations (excluding Europeans) were also assessed through principal components analysis. The first two principal components (PC) accounted for 68.2% of the genetic variance observed and their representation in Figure 3 shows a Pacific population pattern similar to that of the neighbour-joining tree. The first PC underlies the differentiation of Australian and New Guinean populations that is determined mainly by the D1, TPA25, and FXIIIB frequencies (with absolute correlations higher than 70%). The second PC shows the clear separation of the Taiwanese group (Taiwan and Philippines) from Polynesians, with most SE Asian and American populations showing an intermediate position. The RR *Rapanui* sample appears particularly close to Samoa (Polynesia) and also related to Java (*Wallacea*).

A similar picture was obtained from the Delaunay analysis, which imposes the principal genetic boundaries over the map (Figure 4). The first two genetic boundaries split the Australians and Melanesians from all others and from each other. The third boundary distinguished the Taiwanese group, indicating that the *Alu* variation is more consistent with a Polynesian origin from the *Wallacea* region than other models such as the "Fast Train" or the "Entagled Bank".

Discussion

This paper, focusing on the variation of 18 *Alu* insertion polymorphisms in Easter Island, should be considered in the context of other previous studies on the genetic characteristics of this population (such as Cruz-Coke, 1988; Etcheverry, 1967; Hagelberg *et al.* 1994; Hurles *et al.* 2003; Lum *et al.* 1998; Serjeantson *et al.* 1989) but

© 2006 The Authors
Journal compilation © 2006 University College London

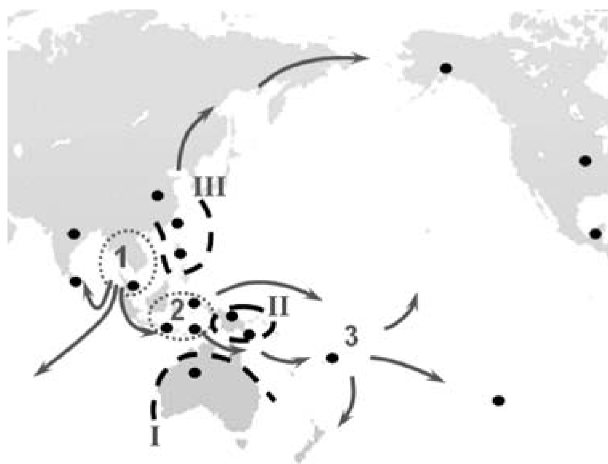

Figure 4 Delaunay triangulation between Pacific populations based on genetic distances from *Alu* insertion polymorphisms. The first (I), second (II), and third (III) most significant boundaries recognized are indicated as dashed lines. Arrows indicate the most suitable scenario for population expansions (1: human arrival to SE Asia between 45,000 and 20,000 years BP; 2: settlement of the Wallacea region through the Holocene; 3: expansion and peopling of Near and Remote Polynesia from 3,500 to 700 years BP).

with two remarkable novelties. This is the first time that a large battery of *Alu* polymorphic elements has been determined in this Pacific population. Furthermore, this is also the first time that two genealogically different samples from the same population have been analyzed jointly. The differentiation between these two samples was clearly established on the basis of personal information, and confirmed by genealogical records available from a demographic study carried out on the same population (Hernández *et al*. 2000). The features of the two Easter Island samples analyzed allowed a direct approach to both the current population heterogeneity and potential changes in the genetic pool of the Island that have occurred during the last century and that are possibly occurring at present.

The autosomal markers examined show the existence of clear genetic differentiation within the current Easter Island population. This variation appears from the simple comparison of marker distributions, and seems to be statistically consistent since differences were actually present in 22.22% of the comparisons, a value clearly exceeding the 5% expected by chance. The genetic distance between the two Easter Island samples (0.051) was equivalent to the average distance among SE Asian and Pacific populations, and more than three times higher than that among Europeans (see Table 3), confirming the diversity within modern Easter Islanders.

According to the features of the samples analyzed this differentiation might only be attributable to the effects of recent (last century) migration. Such an effect is consistent both with the detected increase of gene diversity in the NR sample, and the approximately 50% reduction of genetic distances between this sample and all external population groups as compared with the RR island sample. Also the mutation stability of the *Alu* markers and the relatively short time period (one century) covered by the samples analyzed, points out migration as the most plausible explanation. These results are concordant with recent demographic studies showing an important and progressive increase in the exogamy rate during the 20th century (Hernández *et al*. 2000).

Although the lack of a defined phylogeography for *Alu* markers imposes clear restrictions on our ability to identify the main sources of migration into the island, indirect evidence can be obtained from the admixture analysis we carried out. This analysis shows a European contribution in the NR island sample around twice greater than in the RR *Rapanui* group (m = 0.484 vs. m = 0.296), suggesting that European gene flow might have been one of the most apparent change factors in the *Alu* island gene pool. The introduction of European genes into Polynesia is a well known event, historically documented from the 18th century onward, and has been noted in several Oceanic islands from Y-chromosome variation studies (Hurles *et al*. 2002; Underhill *et al*. 2001). In most cases this was mainly through males and based on small initial numbers, but its genetic effect has been enhanced by particular island conditions (small population sizes, different resistance to epidemics, etc.). In the case of Easter Island the putative European immigration, suggested by the autosomal genetic differentiation between the two current samples, most likely reflects the cumulative effects of initially limited arrivals since the political annexe to Chile in 1888, which increased remarkably by the 1960s with the opening of the island to the rest of the world. According to demographic data, by this time the most important source of migrants was mainland Chile, and hence the arrival of European genes could have been through

© 2006 The Authors
Journal compilation © 2006 University College London

Chilean people with a substantial European genetic component.

Concerning genetic traces on the origin of the *Rapanui* population living in Easter Island around a century ago (represented by the RR sample), the *Alu* polymorphisms indicate a greater affinity to SE Asian populations. The significant apportionment of the genetic variance among geographic/population groups detected indicates that the eight *Alu* insertion polymorphisms used for comparisons have enough power to resolve differences among SE Asian, Taiwanese group, American and European populations; that is, the outside population groups related to Polynesian ancestry in general, and the origin of *Rapanui* islanders in particular. Genetic distance and PC analyses show the similarity of the *Rapanui* to SE Asians and Wallacean populations, as well as their clear differentiation both from Europeans and Australians and Melanesians. The highest genetic affinity with SE Asian populations is concordant with previous studies of globin variants (Hill *et al*. 1989), and some more recent data on the mtDNA and Y chromosome (Kayser *et al*. 2000; Oppenheimer & Richards, 2001), but in contrast with other data suggesting higher affinity of Polynesians with Taiwanese groups (for example initial mtDNA studies: Lum *et al*. 1998; HLA system: Hagelberg *et al*. 1999). Although later differentiation of the Taiwanese group through drift, and other effects, in the past 5,000 years cannot be discarded (Melton *et al*. 1998), our results give overall support to the Polynesian settlement model named 'Voyaging Corridor' or 'Slow Boats to Polynesia' (Oppenheimer & Richards, 2001). This model places Polynesian ancestors among the populations coming from the seafarers in the *Wallacea* region. The expansion of these seafarers would have been taking place before the arrival of the Neolithic in the region. The main tool implicated in these expansions is the canoe, which would have allowed the advance of the expansion wave towards nearby Polynesia as a first step, with subsequent journeys of longer distances towards the remote islands of Polynesia such as *Rapanui*. This model implies an intermediate passage in Coastal Melanesia, with a certain degree of admixture between local populations and Polynesian ancestors. The *Alu* markers are consistent with a gene flow of 15% (data not shown) into the Polynesian gene pool from Coastal, but not from Highland, Melanesia, adding independent support to the model.

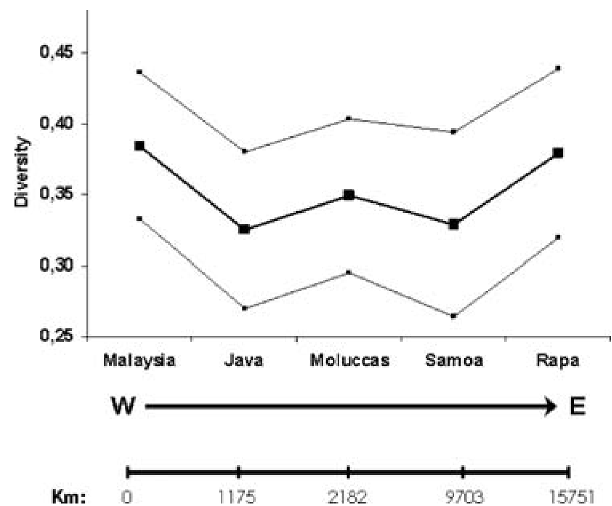

Figure 5 *Alu* diversity among five Pacific populations, ordered according to their geographic position. The thin lines indicate standard errors of diversity measures.

Argued genetic evidence supporting a western ancestry for Polynesians is the west-to-east cline of decreasing diversity throughout the Pacific, which has been described using different loci (Martinson *et al*. 1993). However, when *Alu* marker distributions were compared the *Rapanui* exhibited a remarkably high diversity, which was also detected with other markers (Hurles *et al*. 2003), and is not consistent with its geographical position (Figure. 5) and demographic history. This higher-than-expected diversity among *Rapanui* at the beginning of the 20th century might likely be accounted for by historic events relating to the Peruvian slave trade throughout Polynesia in the second half of the 19th century. This episode implied initially a strong depopulation (by up to 80%) of many islands, included *Rapanui*, and a final disastrous repatriation that cost the lives of numerous Polynesians (Maude, 1981). This repatriation entailed the arrival in *Rapanui* of liberated slaves from different Polynesian islands and of Americans, most likely with a considerable degree of European admixture, immediately prior to the population crash which was mainly caused by epidemics. Although the number of people repatriated was relatively low they had a significant contribution to the succeeding generations of *Rapanui*, due to their relative resistance (as witnessed by their survival

© 2006 The Authors
Journal compilation © 2006 University College London

despite the poor slave working conditions and the voyage) to the disease that caused the epidemic, according to historical records (Maude, 1981; McCall, 1996). The diverse origins of the immediate ancestors of the *Rapanui* at the end of the 19th century explain the relatively high *Alu* diversity observed in the context of Pacific populations. Since most of the repatriated were Polynesians of diverse origins, the genetic affinities shown by the *Alu* markers are revealing both traces of the original human settlers of Polynesia and the effects of historical events that determined the gene pool of the modern *Rapanui*.

With regard to the potential relationships of *Rapanui* with Native Americans (Heyerdahl, 1950), as suggested elsewhere by genetic data (Cann & Lum, 1996) and by minor cultural traits (Green, 2000), the autosomal loci examined failed to shown any special genetic affinity supporting significant prehistoric contacts between the *Rapanui* and those populations. The Native American samples considered here showed closest genetic distances with SE Asian populations, leaving open the question of the different origins of human migrations that settled the American continent.

In summary, our study with *Alu* insertion polymorphisms has shown the existence of genetic heterogeneity in the current *Rapanui* population, illustrating the potential impact that recent migratory events might have had in changing the gene pools of relatively small human groups, such as those living in certain islands. It also emphasizes the need to integrate historical information in analyses to understand the genetic history of human groups, independent of the variation rate of the markers used. Finally, it stresses the importance of accurate sampling for human population studies, to gain insights into the nature and timing of changes that have affected human groups.

Acknowledgements

We would thank the *Rapanui* people, especially all the anonymous islanders who generously decided to participate in the study, and the Hanga Roa Hospital (R Diaz, C Concha, P de la Barra) for their invaluable collaboration. This work was supported by Spanish Ministry grant PB2002-01224; as part of the European Science Foundation EUROCORES Programme OMLL it was also supported by the Ministerio de Ciencia y Tecnologia (BSO2002-10225E) and the EC Sixth Framework Programme under Contract no.ERAS-CT-2003-980409. We wish to thank Prof. Henry Harpending for his helpful comments. E.G.P. was supported by Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya grant 2001FI 00177.

References

Arcot, S. S., Fontius, J. J., Deininger, P. L. & Batzer, M. A. (1995) Identification and analysis of a "young" polymorphic *Alu* element. *Biochim Biophys Acta* **1263**, 99–102.

Arcot, S. S., Adamson, A. W., Lamerdin, J. E., Kanagy, B., Deininger, P. L., Carrano, A. V. & Batzer, M. A. (1996) *Alu* fossil relics – distribution and insertion polymorphism. *Genome Res* **6**, 1084–1092.

Battilana, J., Bonatto, S. L., Freitas, L. B., Hutz, M. H., Weimer, T. A., Callegari-Jacques, S. M., Batzer, M. A., Hill, K., Hurtado, A. M., Tsuneto, L. T., Petzl-Erler, M. L. & Salzano, F. M.(2002) *Alu* insertions versus blood group plus protein genetic variability in four Amerindian populations. *Ann Hum Biol.* **29** (3), 334–347.

Batzer, M. A., Deininger, P. L. (1991) A human-specific subfamily of *Alu* sequences. *Genomics* **9**, 481–487.

Brace, C. L., Nelson, A. R., Seguchi, N., Oe, H., Sering, L., Pan, Q. F., Li, Y. Y. & Tumen, D. (2001) Old World sources of the first New World human inhabitants: A comparative craniofacial view. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10017–10022.

Bertorelle, G. & Excoffier, L. (1998) Inferring admixture proportions from molecular data. *Mol Biol Evol* **15**, 1298–1311.

Brassel, K. E. & Reif, D. (1979) A procedure to generate Thiessen polygons. *Geogr Anal* **11**, 289–303.

Cruz-Coke, R. (1988) Estudios biomédicos chilenos en Isla de Pascua. *Rev Méd Chile* **116**, 818–821.

Cann, R. L. & Lum, J. K. (1996) Mitochondrial myopia: Reply to Bonatto et al. *Am J Hum Genet* **59**, 256–258.

Capelli, C., Wilson, J. F., Richards, M., Stumpf, M. P. H., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V. L., Ko, T. M. & Goldstein, D. B. (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet* **68**, 432–443.

Cavalli-Sforza, L. L. & Feldman, M.W. (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet* Suppl, 266–275.

Diamond, J. (1988) Express train to Polynesia. *Nature* **336**, 307–308.

Dyen, I. (1971) The Austronesian languages and proto-Austronesian. In: *Current trends in linguistics no. 8: Linguistics in Oceania* (ed. T. D. Sebeok), pp 5–54, Mouton, The Hague.

Edwards, M. C. & Gibs, R. A. (1992) A human dimorphism resulting from loss of an *Alu*. *Genomics* **14**, 590–597.

© 2006 The Authors
Journal compilation © 2006 University College London

Etcheverry, R. (1967) Blood groups in natives of Easter Island. *Nature* **216**, 690–691.

Felsenstein, J. (1989) PHYLIP – phylogeny inference package. *Cladistics* **5**, 164–166.

Flenley, J. R. (1993) The paleoecology of Easter Island, and its ecological disaster. In: *Easter Island Studies. Contributions to the history of Rapanui in Memory of William T Mulloy* (ed: S. R. Fisher), pp 27–45, Oxford: Oxford Monograph 32.

Guo, S. & Thomson, E. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361–372.

Green, R. C. (2000) A range of disciplines support a dual origin for bottle gourd in the Pacific. *J Polyn Soc* **109**, 191–197.

González-José, R., González-Martín, A., Hernández, M., Pucciarelli, H. M., Sardi, M., Rosales, A. & Van Der Molen, S. (2003) Craniometric evidence for Palaeoamerican survival in Baja California. *Nature* **425**, 62–65.

Hagelberg, E., Quevedo, S., Turbon, D. & Clegg, J. B. (1994) DNA from Ancient Easter-Islanders. *Nature* **369**, 25–26.

Hagelberg, E., Kayser, M., Nagy, M., Roewer, L., Zimdahl, H., Krawczak, M., Lió, P. & Schiefenhövel, W. (1999) Molecular genetic evidence for the human settlement of the Pacific: analysis of mtDNA, Y chromosome and HLA markers. *Phil Trans R Soc Lond B* **354**, 141–152.

Hernández, M., García-Moro, C., Moral, P. & González-Martín, A. (2000) Population evolution in 20th century Easter Island: Endogamy and admixture. *Hum Biol* **72**, 359–377.

Heyerdahl, T. (1950) Kontiki: across the Pacific by raft. Rand McNally, Chicago.

Hill, A. V., Bowden, D. K., Trent, R. J., Higgs, D. R., Oppenheimer, S. J., Thein, S. L., Mickleson, K. N. P., Weatherall, D. J. & Clegg, J. B. (1985) Melanesians and Polynesians share a unique alpha-thalassaemia mutation. *Am J Hum Genet* **37**, 471–480.

Hill, A. V. S., O'Shaughnessy, D. F. & Clegg, J. B. (1989) Haemoglobyn and globin gene variance in the Pacific. In: *The colonisation of the Pacific: a genetic trail* (eds: A. V. S. Hill, & S. W. Serjeantson), pp 246–285, Clarendon Press, Oxford.

Hurles, M. E., Nicholson, J., Bosch, E., Renfrew, C., Sykes, B. C. & Jobling, M. A. (2002) Y chromosomal evidence for the origins of Oceanic-speaking peoples. *Genetics* **160**, 289–303.

Hurles, M. E., Maund, E., Nicholson, J., Bosch, E., Renfrew, C., Sykes, B. C. & Jobling, M. A. (2003) Native American Y chromosomes in Polynesia: The genetic impact of the Polynesian slave trade. *Am J Hum Genet* **72**, 1282–1287.

Kayser, M., Brauer, S., Weiss, G., Underhill, P. A., Roewer, L., Schiefenhovel, W. & Stoneking, M. (2000) Melanesian origin of Polynesian Y chromosomes. *Current Biology* **10**, 1237–1246.

Lum, J. K., Cann, R. L., Martinson, J. & Jorde, L. B. (1998) Mitochondrial and nuclear genetic relationships among Pacific island and Asian populations. *Am J Hum Genet* **63**, 613–624.

Martinson, J. J., Harding, R. M., Philippon, G., Saintemarie, F. F., Roux, J., Boyce, A. J. & Clegg, J. B. (1993) Demographic reductions and genetic bottlenecks in humans: minisatellite allele distributions in Oceania. *Hum Genet* **91**, 445–450.

Maude, H. E. (1981) *Slavers in paradyse: the peruvian labour trade in Polynesia*. Institute of Pacific Studies, University of the South Pacific, Suva.

McCall, G. (1996) El pasado en el presente Rapanui (Isla de Pascua). In: *Culturas de Chile Vol 2 Etnografia*. (eds: J. Hidalgo et al) Andres Bello, Buenos Aires.

Melton, T., Peterson, R., Redd, A. J., Saha, N., Sofro, A. S. M., Martinson, J. & Stoneking, M. (1995) Polynesian genetic affinities with Southest-Asian populations as identified by mtDNA analysis. *Am J Hum Genet* **57**, 403–414.

Melton, T., Clifford, S., Martinson, J., Batzer, M. A. & Stoneking, M. (1998) Genetic evidence for the Proto-Austranesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am J Hum Genet* **63**, 1807–1823.

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

Novick, G. E., Novick, C. C., Yunis, J., Antunez de Mayolo, P., Douglas Sheer, W., Deininger, P. L., Stoneking, M., York, D. S., Batser, M. A. & Herrera, R. J. (1998) Polymorphic *Alu* insertions and the Asian origin of Native American populations. *Hum Biol* **70**, 23–39.

Oppenheimer, S. & Richards, M. (2001) Fast trains, slow boats, and the ancestry of the Polynesian islanders. *Science Progress* **84**, 157–181.

Pietrusewsky, M. (1997) The people of Ban Chiang: and early Bronze Age site in northeast Thailand. *Indo-Pacific Prehistory Association Bulletin* **16**, 119–147.

Redd, A. J., Takezaki, N., Sherry, S. T., McGarvey, S. T., Sofro, A. S. M. & Stoneking, M. (1995) Evolutionary history of the COII/tRNA(Lys) intergenic 9-base-pais deletion in human mitochondrial DNAs from the Pacific. *Molecular Biology and Evolution* **12**, 604–615.

Reynolds, J., Weir, B. S. & Cockerman, C. C. (1983) Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* **105**, 767–779.

Saitou, N. & Nei, N. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Schneider, S., Roessli, D. & Excoffier, L. (2000) Arlequin ver. 2000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

© 2006 The Authors
Journal compilation © 2006 University College London

Serjeantson, S. W. & Hill, A. V. (1989) The colonisation of the Pacific: the genetic evidence. In: *The colonisation of the Pacific: a genetic trai*. (eds: A. V. S. Hill, & S. W. Serjeantson), pp 286–294, Clarendon Press, Oxford.

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

Sykes, B., Leiboff, A., Lowbeer, J., Tetzner, S. & Richards, M. (1995) The Origins of the Polynesians - An Interpretation from Mitochondrial Lineage Analysis. *Am J Hum Gen* **57**, 1463–1475.

Terrell, J. E. & Welsch, R. L. (1997) *The prehistoric exploration and colonisation of the Pacific*. Cambridge University Press, Cambridge.

Underhill, P. A., Passarino, G., Lin, A. A., Marzuki, S., Oefner, P. J., Cavalli-Sforza, L. L. & Chambers, G. K. (2001) Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Human Mutation* **17**, 271–280.

Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V., Batzer, M. A., Harpending, H. C., Rogers, A. R. & Jorde, L. B. (2001) Patterns of ancestral human diversity: an analysis of *Alu*-insertion and restriction-site polymorphisms. *Am J Hum Genet* **68**, 738–752.

Received: 4 November 2005
Accepted: 23 February 2006

© 2006 The Authors
Journal compilation © 2006 University College London