Journal of Human Genetics (2010) 55, 827–833

www.nature.com/jhg

ORIGINAL ARTICLE

Mixed origin of the current Tunisian population from the analysis of *Alu* and *Alu*/STR compound systems

Wifak El Moncer[1], Esther Esteban[2], Raoudha Bahri[1], Magdalena Gayà-Vidal[2], Robert Carreras-Torres[2], Georgios Athanasiadis[2], Pedro Moral[2] and Hassen Chaabani[1]

During successive historical periods, Tunisia has been a crossroads of multiple civilizations and their corresponding key population movements. The aim of this study was to provide genetic information relating to the mixed origin of the Tunisian population, and to analyze its genetic relationship with other North African and Mediterranean populations. A set of 16 *Alu* and 3 *Alu*/STR compound systems has been analyzed in 268 autochthonous Tunisians from the north-center and the south of the country. Our two sampled populations showed no significant differentiation from one another in any of the three *Alu*/STR compound systems, whereas the analysis of the 16 *Alu* markers revealed a significant genetic differentiation between them. A sub-Saharan component shown by the three *Alu*/STR combinations is more noticeable in our north-center sample than in that of the south. The presence of two *Alu*/STR combinations specific to North African ancestral populations also suggests that the ancient Berber component is relatively more substantial in the north and center regions than in the south. Our Tunisian samples cluster together with other Berber samples from Morocco and Algeria, underpinning the genetic similarity among North Africans regardless of their current linguistic status (Berber or Arabic).
Journal of Human Genetics (2010) **55,** 827–833; doi:10.1038/jhg.2010.120; published online 30 September 2010

Keywords: *Alu* insertion polymorphisms; *Alu*-STR haplotypes; human populations; Tunisians

INTRODUCTION

Tunisia, situated in the northeast extreme of Africa, is open to the Mediterranean Sea all along its coast. The geographical location in the heart of the Mediterranean has allowed Tunisia to have an important role in the history of this region. The Neolithic Age was present in North Africa through three cultural periods.[1] The first, *Neolithic with Sudan origins*, started about 9000 years before present (YBP) in the extreme south of North Africa, where the current Saharan regions of Algeria are, and is characterized by an ethnic contribution from Sudan. The second, *Neolithic with Capsian origins*, started about 7000 YBP, represents the continuity of the local Capsian culture (10 000–8000 YBP) and is considered to be the source of 'proto-Mediterranean' peoples. The third, *Mediterranean Neolithic*, embraces the northern coastal regions and very likely represents the continuity of the local Oranean culture (17 000–10 000 YBP), together with influences from some civilizations from the northern side of the Mediterranean.

Historians name 'Berbers' the people living in North Africa in the last 6000 years. This population represents, at least in part, the descendants of the aforementioned ancient peoples. As a result, the current Tunisian population is probably composed by an ancient Berber background together with influences from the different civilizations settled in this region in historical times: Phoenicians from Tyre (the present-day Lebanon), who founded the celebrated city of Carthage, Romans, Vandals, Byzantines and, finally, the substantial expansion of Arabs in the seventh century AD. During all these periods, local Berbers were opposed to the invaders, who were unable to dominate all Tunisian regions, with the exception of the Arabs, who settled permanently in Tunisia, as well as in other North African countries.[2] In fact, although at first the Arabs met remarkable resistance from the Berbers, they persuaded them to adopt Islam, to learn the Arabic language and to accept intermarriages. Finally, Tunisia received additional, less important contributions, such as that of the Ottoman Turks during the sixteenth century.

In spite of the general adoption of the Muslim Arab culture, some large Berber groups in Algeria and Morocco have kept their Berber language and customs until now and avoided intermarriage. All the same, the original Berber and Arab populations in Tunisia were widely mixed except for few small Berber communities, more or less geographically isolated, such as Takrouna and Jeradou in the center; Douirete-Chenini in the extreme south; and a group living in the 'Gallala' region of Djerba island.[3,4] These small Berber groups, often comprising less than 5000 individuals, can be considered to be isolates, supported by several genetic studies showing a remarkable genetic heterogeneity among them.[3–7] Small effective sizes, founder effects and variation of sub-Saharan African traces in their gene pool could be the

[1]Laboratory of Human Genetics and Anthropology, Faculty of Pharmacy, University of Monastir, Monastir, Tunisia and [2]Secció d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Barcelona, Spain
Correspondence: Dr E Esteban, Secció d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Avinguda Diagonal, 645, Barcelona 08028, Spain.
E-mail: mesteban@ub.edu
Received 20 May 2010; revised 2 September 2010; accepted 4 September 2010; published online 30 September 2010

1. North Spain: NSpain (114)
2. Pas Valley: PSpain (99)
3. Basque country: BasqueC (110)
4. Central Spain: CSpain (120)
5. South Spain: SSpain (100)
6. South France: SFrance (55)
7. Majorca: Majorca (114)
8. Corsica Centre: CorC (102)
9. Corsica West Coast: CorWC (96)
10. Sardinia Centre: SarC (100)
11. Sardinia WC: SarWC (98)
12. Sicily West: Sicw (100)
13. Sicily East: SicE (102)
14. Greece: Greece (96)
15. Turkey: Turkey (96)
16. Egyptian Siwa Berbers: SiwaB (86)
17. Tunisia North-Centre: TunNC (120)
18. Tunisia South: TunS (148)
19. Algerian Mozabite Berbers: MzabB (109)
20. High Atlas Berbers Asni: AsniB (119)
21. High Atlas Berbers Amizmiz: AmizB (86)
22. Middle Atlas Berbers Khenifra: KheniB (136)
23. North East Atlas Berbers Bouhria: BouhB (111)
24. Arab-speaking Moroccans: ArabM (98)

Figure 1 Geographical location of the samples studied and other populations used for comparisons.

main factors contributing to current heterogeneity of these Berber groups. Nevertheless, leaving these particular groups aside, the general Tunisian population shows a considerable degree of genetic homogeneity.[8]

Bearing in mind the historical data quoted, it could be said that the present-day Tunisian population has a uniform Muslim Arab culture, but in genetic terms it represents a mixture mainly composed by Berbers—the autochthonous population—and Arabs, with a relatively small contribution from other surrounding peoples that ruled North Africa in later periods. The aim of this study was to determine to what extent current Tunisians show traces of these mixed origins, taking advantage of the combined information derived from autosomal *Alu*/STR compound systems. These particular markers have been selected for two main reasons. The first is the widely contrasted informative nature of the *Alu* polymorphisms[9] because of their stability, low mutation rate and known ancestral state. The second is the high degree of qualitative information provided by an STR linked to an *Alu* marker. Information from haplotype *Alu*/STR frequencies has been successfully used to estimate fine genetic relationships in the Mediterranean region.[10–12]

The main question of this work will be addressed through some specific objectives: (1) determining the degree of heterogeneity of the general Tunisian population through the comparison of two samples from different geographical areas; (2) analyzing the genetic relationships of these samples in a wider North African and Mediterranean context, in particular with those groups historically related to the Tunisian population and (3) deducing more about the mixed origin of the current general Tunisian population, particularly from the analysis of *Alu*/STR compound systems.

MATERIALS AND METHODS

Blood samples of 268 autochthonous Tunisian individuals were collected: 120 from north-center regions and 148 from the south. All individuals were healthy, unrelated donors, who signed an informed consent approved by the ethical committees of the Universities involved in the study. We have considered the north and the center in a unique north-center sample because preliminary results[8] showed a high degree of homogeneity between these samples. The south of Tunisia, particularly the extreme south, is compared for the first time with the rest of the Tunisian population. The subjects from the southern sample are from the extreme southern regions of Gabès, Gbelli and Mednine.

DNA from all individuals has been genotyped for 16 *Alu* polymorphisms (DM (chromosome 19), HS4.69 (chromosome 6), HS4.32 (chromosome 12), Ya5NBC221 (chromosome 22), Sb19.3 (chromosome 19), HS2.43 (chromosome 1), Sb19.12 (chromosome 19), Yb8NBC120 (chromosome 22), Yb8NBC125 (chromosome 22), PV92 (chromosome 16), FXIIIB (chromosome 1), A25 (chromosome 8), CD4 (chromosome 12), TPA25 (chromosome 8), APOA1 (chromosome 11), ACE (chromosome 17)) and three STRs from the CD4, FXIIIB and DM loci. Supplementary Table 1 includes the GenBank/EMBL accession number of each *Alu*, together with information about the frequency population ranges that have been established using populations of Figure 1.

Genomic DNA was extracted from blood by standard phenol-chloroform techniques. *Alu* genotyping was carried out by PCR followed by electrophoresis separation on 2% agarose gels. Three STRs, a pentanucleotide from the CD4 locus, a tetranucleotide from the FXIIIB gene and a trinucleotide from the DM locus, were determined by PCR amplification with fluorescent-labeled primers. PCR products were electrophoresed on ABI PRISM 3700 DNA sequencer (Applied Biosystems, Foster City, CA, USA). GeneScan and GeneMapper 3.0 programs (ABI PRISM; Applied Biosystems) were used to genotype individuals. Technical references on PCR and electrophoresis are extensively explained in our previous works[11,12] for both *Alu* and *Alu*/STR combinations. The nomenclature of *Alu*/STR combinations consists of a number indicating the size in base pairs of the corresponding STR allele, followed by a symbol + (presence of the *Alu* element) or − (absence of *Alu*).

Allele frequencies were calculated by direct counting and Hardy–Weinberg equilibrium was checked through an exact test.[13] Heterozygosity by population and by locus was estimated according to Nei's formula.[14] Maximum likelihood haplotype frequencies were computed using the EM algorithm. The geographical structure of the allele frequency variance was tested by hierarchical analyses of molecular variation using Wright's F statistics from populations clustered according to geographical criteria. These calculations were performed using the GenePop 3.3 (GenePop, Montpellier, France) and Arlequin packages (Arlequin, Berne, Switzerland).[15,16] Population genetic relationships for the *Alu* and *Alu*/STR data were also assessed by pairwise F_{ST} genetic distances,[17] and represented by a multidimensional scaling (MDS) plot from the distance matrix. Levels of genetic admixture were estimated with the Leadmix program (Leadmix, London, UK).[18] This Fortran program has been designed to obtain maximum likelihood estimates and 95% confidence intervals of several parameters of

Table 1 *Alu* insertion frequencies and heterozygosity values in Tunisia north-center (Tunisia NC) and Tunisia south (Tunisia S)

	Tunisia NC	*Tunisia S*		*Tunisia NC*	*Tunisia S*		*Tunisia NC*	*Tunisia S*
DM			*Sb19.12*			*A25*		
N	82	128	*N*	106	133	*N*	118	116
Insertion	0.335	0.437	Insertion	0.207	0.244	Insertion	0.123	0.121
Heterozygosity	0.448	0.494	Heterozygosity	0.330	0.371	Heterozygosity	0.216	0.213
HS4.69			*Yb8NBC120*			*CD4*		
N	100	114	*N*	115	145	Heterozygosity	113	138
Insertion	0.360	0.338	Insertion	0.452	0.476	Insertion	0.699	0.739
Heterozygosity	0.463	0.449	Heterozygosity	0.498	0.501	Heterozygosity	0.423	0.387
HS4.32			*Yb8NBC125*			*TPA25*		
N	81	114	*N*	112	146	*N*	113	125
Insertion	0.741	0.675	Insertion	0.129	0.130	Insertion	0.584	0.524
Heterozygosity	0.386	0.440	Heterozygosity	0.226	0.227	Heterozygosity	0.488	0.501
Ya5NBC221			*PV92*			*APOA1*		
N	116	143	*N*	103	101	*N*	118	127
Insertion	0.901	0.937	Insertion	0.359	0.233	Insertion	0.902	0.823
Heterozygosity	0.179	0.118	Heterozygosity	0.462	0.359	Heterozygosity	0.177	0.293
Sb19.3			*FXIIIB*			*ACE*		
N	115	146	*N*	115	142	*N*	119	144
Insertion	0.752	0.709	Insertion	0.304	0.324	Insertion	0.281	0.264
Heterozygosity	0.374	0.414	Heterozygosity	0.425	0.440	Heterozygosity	0.406	0.390
HS2.43								
N	87	113						
Insertion	0.0345	0.093						
Heterozygosity	0.067	0.169						

Information about population ranges of each *Alu*, established using populations of Figure 1, can be checked in Supplementary Table 1.

ancestral, parental and hybrid populations, together with time estimations since the split of populations. The method and program apply to the case of more than two parental populations contributing to the admixture. The samples can be analyzed for different markers (for example, DNA sequence, microsatellites) and are used to estimate admixture proportions and genetic drift.

For comparative purposes, 22 Mediterranean samples previously tested for the same genetic markers, and a sub-Saharan African sample from the Ivory Coast (95 sampled individuals) were selected.[11,12] The geographical location and sample sizes of the Mediterranean populations used in this study are indicated in Figure 1. The samples included eight continental north Mediterranean groups spanning from the Iberian Peninsula to Turkey; seven samples from the western Mediterranean islands of Majorca, Corsica, Sardinia and Sicily; and, finally, a set of seven North African populations composed by six Berber groups and one Arab-speaking Moroccan sample. Population ranges and pairwise population comparisons have been established using the complete set of 22 samples, but for genetic distance and analysis of molecular variance calculations some samples (Spanish Pas Valley, Sardinians and Siwa Egyptians) were excluded due to their extreme differentiation, discussed in previous reports.[11,12]

RESULTS

Alu and *Alu*/STR genetic variability in north-center and south Tunisian samples

The *Alu* allele frequencies for the 16 loci examined are shown in Table 1. All loci fit Hardy–Weinberg equilibrium after Bonferroni correction except for *Alu* HS4.32 and Ya5NBC221 in the southern Tunisian sample.

On average, the gene diversity value observed in the north-center (0.360 ± 0.121) Tunisian sample was slightly higher than that of the south (0.348 ± 0.133). *Alu* frequency comparisons, checked through the exact test of population differentiation, yielded significant differences between our two samples for DM ($P=0.035$), HS2.43 ($P=0.028$), PV92 ($P=0.006$) and APOA1 ($P=0.013$), and also across the 16 markers considered ($P=0.011$). As to population variation ranges (Supplementary Table 1) established through data from 22 Mediterranean populations, our populations occupied intermediate positions in all cases except south Tunisia for *Alu* APOA1, which showed the lowest value. Pairwise population differences revealed a remarkable degree of genetic heterogeneity in the whole Mediterranean context. Across the 16 loci, the exact test indicated that 260 comparisons of 276 were significant. The remaining 16 nonsignificant population comparisons included, among others, the comparisons of the north-center Tunisian sample with both northeast Atlas and Middle Atlas Moroccan Berbers.

Alu/STR haplotype frequencies are shown in Table 2 for the CD4, FXIIIB and DM loci. In all three compound systems, *Alu* and STR alleles were in linkage disequilibrium. Haplotype diversity values in south Tunisia for CD4 (0.809 ± 0.013) and DM (0.765 ± 0.023) were slightly higher than those reported for the north-center sample (0.836 ± 0.016 and 0.780 ± 0.026, respectively). The opposite trend was observed for the FXIIIB haplotypes (Table 2). In agreement with that, the number of different haplotypes (only those having frequency values $>1\%$) in the north-center sample was slightly higher than that in south Tunisia for the CD4 (13 vs 9 haplotypes, respectively) and DM loci (9 vs 8), but not for F13 B (8 vs 9). Population differentiation was considerable in all three systems; from 276 pairs of comparisons,

Table 2 CD4, FXIIIB and DM *Alu*/STR haplotype frequencies in Tunisia north-centre (*N*=120) and Tunisia south (*N*=147)

	Tunisia NC	*Tunisia S*		*Tunisia NC*	*Tunisia S*		*Tunisia NC*	*Tunisia S*
CD4			*FXIIIB*			*DM*		
85(+)	0.226	0.221	172(+)	0.011	0.011	77(+)	0.196	0.237
90(+)	0.060	0.125	180(+)	0.159	0.134	98(+)	0.106	0.147
100(+)	0.015	0.007	184(+)	0.098	0.126	101(+)	0.012	0.006
105(+)	0.016	0.012	188(+)	0.040	0.050	128(+)	0.000	0.012
110(+)	0.235	0.254	172(−)	0.261	0.174	77(−)	0.054	0.023
115(+)	0.096	0.076	176(−)	0.009	0.051	95(−)	0.086	0.051
120(+)	0.005	0.024	180(−)	0.051	0.097	98(−)	0.382	0.383
130(+)	0.015	0.008	184(−)	0.137	0.156	101(−)	0.098	0.077
85(−)	0.045	0.010	188(−)	0.221	0.193	104(−)	0.025	0.028
90(−)	0.201	0.230				107(−)	0.015	0.004
95(−)	0.011	0.000						
110(−)	0.020	0.016						
115(−)	0.010	0.000						
Gene Diversity	0.836 ± 0.016	0.809 ± 0.013		0.826 ± 0.014	0.865 ± 0.008		0.780 ± 0.026	0.765 ± 0.023

STR alleles are expressed in base pairs, the presence of the *Alu* insertion corresponds to the +symbol. Only haplotypes whose frequencies reached values up to 1% are included in the Table.

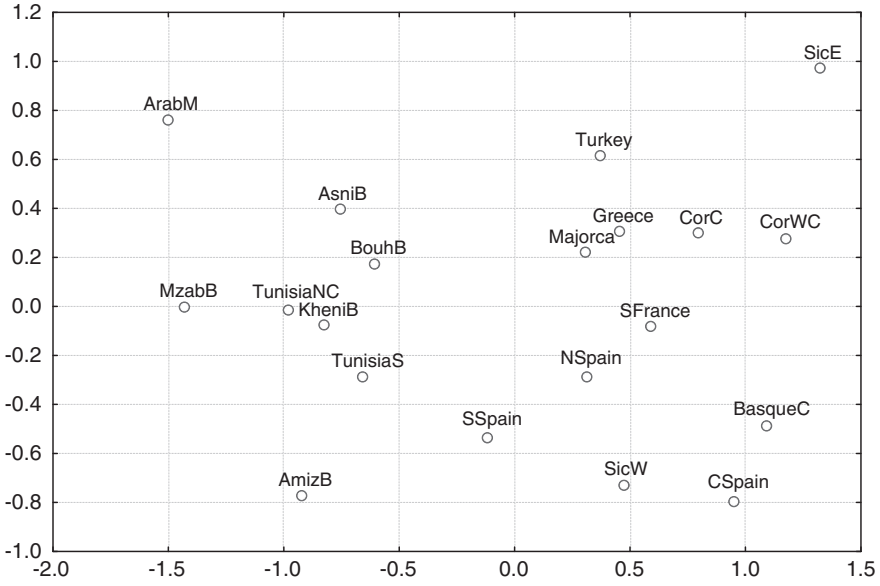

Figure 2 Multidimensional scaling plot based on Reynolds distance matrix, data from 16 *Alus* in 20 populations. Stress of 0.117.

75 for CD4, and 122 for both FXIIIB and DM were statistically significant. However, our Tunisian samples did not show any significant differences between them in any of the three compound systems.

Population relationships in the Mediterranean

Overall genetic relationships among our Tunisian samples and other Mediterranean populations were assessed through F_{ST}-related genetic distances (depicted in MDS plots) for 16 *Alu* polymorphisms (Figure 2) and for 3 *Alu*/STR compound systems (Figure 3). In both matrices, distance values were significantly different from zero in more than 85% of the cases. The MDS representation of *Alu*-based genetic distances underlined the difference between the North African samples and the remaining ones, clustering together continental European Mediterraneans and the samples from Majorca, Corsica and Sicily islands. The genetic heterogeneity inside these two groups, estimated by a nonhierarchical analysis of molecular variance, showed similar values (average F_{ST} in North Africa of 1.2%, $P<0.001$; average F_{ST} in European Mediterraneans 1.4%, $P<0.001$). When we compared the genetic structure between North and South we found that, although presenting moderate values, average F_{CT} between these two groups (2%, $P<0.001$) was higher than the average F_{SC} inside each group (1.3%, $P<0.001$).

The plot (Figure 3) based on *Alu*/STR compound systems revealed a similar pattern of population relationships with that found for the *Alu* markers. However, the hierarchical analyses of molecular variance indicated a considerably higher effect of the geographic structure as average F_{CT} between groups (2.6%, $P<0.001$) accounted for more than four times the average F_{SC} within groups (0.6%, $P<0.001$). In the second dimension of the plot, the particular position of our Tunisian samples and the Middle Atlas Berbers probably underpins the genetic differentiation between north and south Mediterraneans.

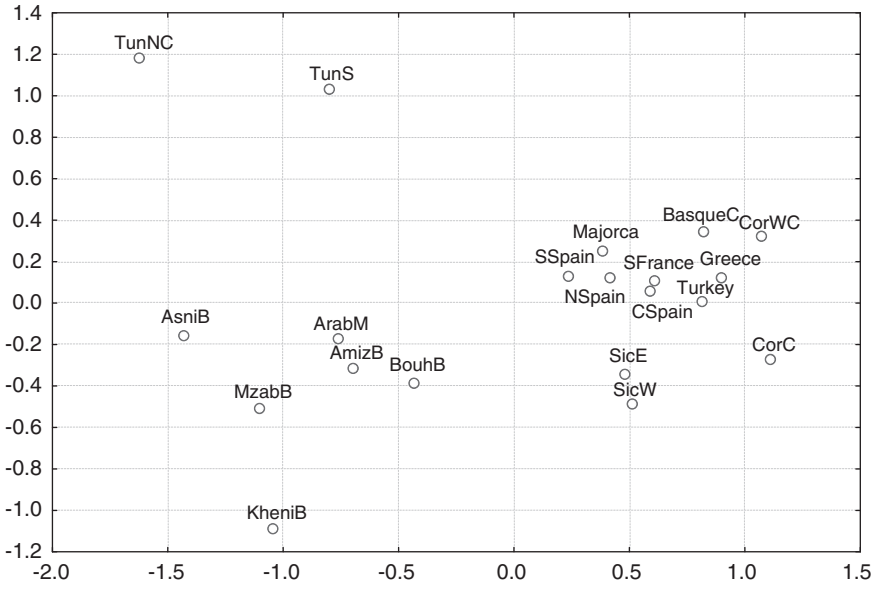

Figure 3 Multidimensional scaling plot based on Reynolds distance matrix, data from CD4, F13B and DM *Alu*/STR haplotypes in 20 populations. Stress of 0.087.

Sub-Saharan African and Berber contribution to the Tunisian genetic background

As stated in previous anthropological studies,[11,19] three CD4 *Alu*/STR haplotypes (100(+), 85(−) and 115(−)) present a clear sub-Saharan African origin. In fact, they are absent or present in negligible frequencies in Europeans and Asians, whereas they are often present in high frequencies in sub-Saharan African groups and in relatively low frequencies in some of their neighboring populations. These haplotypes are also present in our two Tunisian samples with a remarkable quantitative difference: the total number of different sub-Saharan African haplotypes in the north-center sample was four times higher (7%) than that of the south (1.7%). This difference was statistically significant (Fisher's exact probability=0.002). Moreover, two Mediterranean-specific combinations, particularly in the North Africans[11,12] (the CD4 haplotype 110(−) and the DM haplotype 107(−)) also showed higher frequencies in the north-center sample (2 vs 1% for the CD4 110(−) combination; 1.5 vs 0.4% for the DM 107(−) combination) than in the south.

Sub-Saharan African gene flow in Tunisians was tested through LEADMIX simulations under different parental groups. The only consistent results were those based on *Alu*/STR haplotypes taking as parental populations a sub-Saharan African sample on one hand, and a sample from continental south Europeans (described in the Materials and methods section) on the other. For both Tunisian samples, the overall sub-Saharan African contribution reached a similar value: 0.398 (−95% CI 0.228; +95% CI 0.617) for north-center Tunisia and 0.392 (−95% CI 0.190; +95% CI 0.632) for south Tunisia.

DISCUSSION

Recent genetic studies dealing with uniparental[20,21] data have emphasized the genetic distinctiveness of the Mediterraneans, the heterogeneity of some populations from this region due to particular demographic histories and the differential sub-Saharan African gene flow received by both southern and northern shores. Concerning autosomal data,[11,12] the usefulness of the combined use of *Alu* and STR markers to detect fine population relationships has been clearly shown through the study of a wide population set of Mediterranean samples. However, some particular questions still remain to be addressed: the genetic heterogeneity in the Mediterranean seems to be closely related to the differentiation among Berber groups. In that respect, some typical sub-Saharan and Mediterranean *Alu*/STR combinations have been detected with noticeable frequencies in different Berber communities in Morocco and Algeria, but little is known about their presence in the rest of the present-day North African populations, particularly the Tunisian population having a uniform Arab Muslim culture (100% Arab speakers). Do Berber speakers and Arab speakers in the Maghreb share a common genetic background? Should we be talking about a 'Berber' or maybe a 'North African' genetic distinctiveness?

The 16 *Alu* markers and 3 *Alu*/STR compound systems considered here are the first to be jointly described in Tunisian samples from different geographical locations in the attempt to contribute to a better knowledge of the general Tunisian population, allowing us to address the questions stated above.

North-center and south Tunisians showed similar levels of gene diversity for both *Alu* (0.360 and 0.348 for north-center and south, respectively) and *Alu*/STR compound systems (0.814 and 0.813, respectively). These noticeable values are in agreement with those described in other Mediterraneans.[11,12] Genetic differentiation inside Tunisia was significant for *Alu* loci (P=0.011) but not for *Alu*/STR compound systems. Such discrepancy of results could be attributed to the effect of the different number of markers compared (16 vs 3). However, the nature of the markers involved, with remarkably different mutation rates, could also provide an explanation for the above observations: previous studies[11,12] have proposed that *Alu* loci are more suitable to detect ancient relationships, whereas *Alu*/STR haplotypes perform better in quantifying ancestral genes or gene flow thanks to some population specific combinations. Thus, the presence of Berber and sub-Saharan African-specific combinations in remarkably higher frequencies (10.5%) in north-center Tunisia, as compared with the southern sample (3.1%), suggests a certain degree of genetic heterogeneity also for the *Alu*/STR data.

The presence of a sub-Saharan component in the gene pool of the Tunisians was first shown by the GM and immunoglobulin Cγ gene

polymorphisms.[22] The autosomal markers analyzed here have allowed the quantification of sub-Saharan gene flow for the *Alu*/STR haplotypes. Sub-Saharan African contribution in our samples reached 39%. This value is comparable to, and even slightly higher than, other gene flow estimations previously described[11] in several North African populations ranging from 16.8% in Moroccan north-east Atlas Berbers to 37.7% in Mozabite Berbers from Algeria. The presence of noticeable sub-Saharan African traces in present-day Tunisians is in agreement with mtDNA data[23] reporting a higher number of sub-Saharan L lineages in Tunisia (48%) as compared with Morocco (25%).

The qualitative information provided by some particular *Alu*/STR combinations of the CD4 locus, such as 100(+), 85(−) and 115(−), could be another indication of sub-Saharan gene flow. In this case, north-center Tunisia attained a value (7%) considerably higher than that observed in the south of the country (1.7%). These frequencies range from 2.9% in northeast Atlas to 12.3% in Middle Atlas Moroccan Berbers, but they have also been found in Algerian Mozabites (5.8%). The observed fluctuations of sub-Saharan gene flow in North Africa could be related to particular demographic events that may have enhanced the effect of genetic drift on a single locus. Whatever the case, the existence of trans-Saharan African gene flow through the Maghreb is obvious, and has been reported by other genetic studies,[12,23,24] as well as in archeological and historical records.[1]

Notwithstanding, it is important to ask whether this sub-Saharan gene flow is relatively recent or more ancient. Our results are compatible with the latter alternative. In fact, as mentioned above, we have found that the presence of three sub-Saharan Africa-specific CD4 *Alu*/STR combinations is considerably higher in the north-center Tunisian sample than in the one from the south. If the corresponding gene flow occurred in relatively recent times, we should find the opposite trend, because the south of Tunisia would naturally be the first to receive population movements from sub-Saharan countries. Moreover, about 5000 YBP, the immense Sahara desert already had the current severe climate that represents a considerable barrier to human migration, but it was more accessible to human migration[25] before, due to a better climatic conditions. All these data considered together suggest that the sub-Saharan component found in Tunisia is rather ancient and could be traced back to the first stage of Neolithic Age (around 9000 YBP), characterized by an ethnic contribution from present-day Sudan.

When other Mediterranean samples were added to the comparisons, the distance matrices of the *Alu* and *Alu*/STR systems were positively correlated ($P<0.001$, Mantel test), emphasizing the consistency of the relationship pattern. In both cases, the first dimension classified populations into North Africans and Europeans, continental and insular samples included. This geographical differentiation was also evidenced by the overall values of the hierarchical analyses of molecular variance. The genetic variance attributable to the North–South differentiation was significant (*Alu:* F_{SC} within groups=1.3%, F_{CT} between groups=2%, $P<0.001$; *Alu*/STR: F_{SC} within groups=0.6%, F_{CT} between groups=2.6%, $P<0.001$). Our Tunisian samples did not show any close genetic affinity with either the Sicilians or the Turks, two Mediterranean populations that had historical ties with Tunisia.

The Tunisian samples cluster together with the Berber groups from Morocco and Algeria, in agreement with recent works based on other genetic data.[26–28] The close genetic relationship of the two Arab-speaking populations with the Berber-speaking samples could be explained assuming a small number of Arabs coming from the Arabian Peninsula, as compared with that of the autochthonous Berbers, resulting in a weak Arab genetic influence in the current mixed North Africans.

In conclusion, the results discussed here allow us to postulate that the general ancient genetic profile of the native North Africans—the Berbers—is not very different from that of the present-day North African populations, despite some admixture with other peoples, particularly Arabs, during successive historical periods. The populations of the Maghreb seem to share a substantial genetic background, regardless of culture and geography.

ACKNOWLEDGEMENTS

We thank all the anonymous people for their participation in the study. This research has been supported in part by the Tunisian Ministry of Higher Education within the research unity 05/UR/09-04, by the Spanish Ministry of Educación y Ciencia Grant CGL2008-03955, the Generalitat de Catalunya 2009SGR1408 grant, and by the Universitat de Barcelona grants of the Oficina de Mobilitat i Programes Internacionals and the Vicerrectorat de Relacions Internacionals i Institucionals.

1 Gragueb, A. & Mtimet, A. *La Préhistoire en Tunisie et au Maghreb* (Tunis Édition ALIF: Tunisia, 1989).
2 Buresi, P. *Géo-histoire de l'islam* (Belin-Sup.Histoire: Paris, France, 2005)..
3 Chaabani, H., Helal, A. N., van Loghem, E., Langaney, A., Benammar Elgaaied, A., Rivat Peran, L. *et al.* Genetic study of Tunisian Berbers. I. Gm, Am and Km immunoglobulin allotypes and ABO blood groups. *J. Immunogenet.* **11,** 107–113 (1984).
4 Khodjet-el-khil, H., Fadhlaoui-zid, K., Gusmao, L., Alves, C., Benammar-Elgaaied, A. & Amorim, A. Substructure of a Tunisian Berber population as inferred from 15 autosomal short tandem repeat loci. *Hum. Biol.* **80,** 435–448 (2008).
5 Ennafaa, H., Amor, M. B., Yacoubi-Loueslati, B., Khodjet el-khil, H., Gonzalez-Perez, E., Moral, P. *et al.* Alu polymorphisms in Jerba Island population (Tunisia): comparative study in Arab and Berber groups. *Ann. Hum. Biol.* **33,** 634–640 (2006).
6 Frigi, S., Pereira, F., Pereira, L., Yacoubi, B., Gusmão, L., Alves, C. *et al.* Data for Y-chromosome haplotypes defined by 17 STRs (AmpFLSTR Yfiler) in two Tunisian Berber communities. *Forensic. Sci. Int.* **27,** 80–83 (2006).
7 Chaabani, H. & Cox, D. W. Genetic characterisation and origin of Tunisian Berbers. *Hum. Hered.* **38,** 308–316 (1988).
8 Bahri, R., Esteban, E., Moral, P. & Chaabani, H. New insights into the genetic history of Tunisians: data from Alu insertion and apolipoprotein E gene polymorphisms. *Ann. Hum. Biol.* **35,** 22–33 (2008).
9 Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V., Batzer, M. A. *et al.* Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* **68,** 738–752 (2001).
10 Flores, C., Maca-Meyer, N., González, A. M. & Cabrera, V. M. Northwest African distribution of the CD4 *Alu*/microsatellite haplotypes. *Ann. Hum. Genet.* **64,** 321–327 (2000).
11 González-Pérez, E., Moral, P., Via, M., Vona, G., Varesi, L., Santamaria, J. *et al.* 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 (2007).
12 González-Pérez, E., Esteban, E., Via, M., Gaya-Vidal, M., Athanasiadis, G., Dugoujon, J. M. *et al.* Population relationships in the Mediterranean revealed by autosomal genetic data (*Alu* and *Alu*/STR compound systems). *Am. J. Phys. Anthrop.* **141,** 430–439 (2010).
13 Guo, S. & Thomson, E. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48,** 361–372 (1992).
14 Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press: New York, 1987).
15 Raymond, M. & Rousset, F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity* **86,** 248–249 (1995).
16 Excoffier, L., Laval, G. & Schneider, S. ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinf. Online* **1,** 47–50 (2005).
17 Reynolds, J., Weir, B. S. & Cockerman, C. C. Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* **105,** 767–779 (1983).
18 Wang, J. Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics* **164,** 747–765 (2003).
19 Tishkoff, S. A., Dietzsch, E., Speed, W., Pakstis, A. J., Kidd, J. R., Cheung, K. *et al.* Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271,** 1380–1387 (1996).
20 Plaza, S., Calafell, F., Helal, A., Bouzerna, N., Lefranc, G., Bertranpetit, J. *et al.* Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann. Hum. Genet.* **67,** 312–328 (2003).

21 Cruciani, F., La Fratta, R., Trombetta, B., Santolamazza, P., Sellitto, D., Colomb, E B. *et al*. Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol. Biol. Evol.* **24,** 1300–1311 (2007).
22 Chaabani, H., Bech-Hansen, N. T. & Cox, D. W. Restriction fragment length polymorphisms associated with immunoglobulin heavy chain gamma genes in Tunisians. *Hum. Genet.* **73,** 110–113 (1989).
23 Turchi, C., Buscemi, L., Giacchino, E., Onofri, V., Fendt, L., Parson, W. *et al.* Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data. *Forensic Sci. Int. Genet.* **3,** 166–172 (2009).
24 Arredi, B., Poloni, E. S., Parachini, S., Zerjal, T., Fathallah, D. M., Makfrelouf, M. *et al.* A predominantly Neolithic origin for Y-chromosome variation in North Africa. *Am. J. Hum. Genet.* **75,** 338–345 (2004).
25 Pons, A., de Beaulieu, J. L., Guiot, J. & Reille, M. Le pollen remonte le temps climatique. *La recherche* **213,** 518–520 (1991).
26 Ennafaa, H., Cabrera, V. M., Abu-Amero, K. K., González, A. M., B Amor, M., Bouhaha, R. *et al.* Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet.* **10,** 8 (2009).
27 Cherni, L., Fernandes, V., Pereira, J. B., Costa, M. D., Goios, A., Frigi, S. *et al.* Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. *Am. J. Phys. Anthrop.* **139,** 253–260 (2009).
28 Coudray, C., Guitard, E., Kandil, M., Harich, N., Melhaoui, M., Baali, A. *et al.* Study of GM immunoglobulin allotypic system in Berbers and Arabs from Morocco. *Am. J. Hum. Biol.* **18,** 23–34 (2006).
29 Mahadevan, M. S., Amemiya, C., Jansen, G., Sabourin, L., Baird, S., Neville, C. E. *et al.* Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. *Hum. Mol. Genet.* **2,** 299–304 (1993).
30 Arcot, S. S., Adamson, A. W., Lamerdin, J. E., Kanagy, B., Deininger, P. L., Carrano, A. V. *et al.* Alu fossil relics—distribution and insertion polymorphism. *Genome Res.* **6,** 1084–1092 (1996).
31 Carroll, M. L., Roy-Engel, A. M., Nguyen, S. V., Salem, A. H., Vogel, E., Vincent, B. *et al.* Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J. Mol. Biol.* **311,** 17–40 (2001).
32 Arcot, S. S., Adamson, A. W., Risch, G. W., LaFleur, J., Robichaux, M. B., Lamerdin, J. E. *et al.* High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. *J. Mol. Biol.* **281,** 843–856 (1998).
33 Arcot, S. S., Wang, Z., Weber, J. L., Deininger, P. L. & Batzer, M. A. Alu repeats: a source for the genesis of primate microsatellites. *Genomics* **29,** 136–144 (1995).
34 Batzer, M. A., Arcot, S. S., Phinney, J. W., Alegria-Hartman, M., Kass, D. H., Milligan, S. M. *et al.* Genetic variation of recent Alu insertions in human populations. *J. Mol. Evol.* **42,** 22–29 (1996).
35 Edwards, M. C. & Gibbs, R. A. A human dimosphism resulting from loss of an Alu. *Genomics* **14,** 590–597 (1992).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)