Female Gene Pools of Berber and Arab Neighboring Communities in Central Tunisia: Microstructure of mtDNA Variation in North Africa

LOTFI CHERNI,[1] BESMA YAACOUBI LOUESLATI,[1] LUÍSA PEREIRA,[2] HAJER ENNAFAÂ,[1] ANTÓNIO AMORIM,[2,3] AND AMEL BEN AMMAR EL GAAIED[1]

Abstract North African populations are considered genetically closer to Eurasians than to sub-Saharans. However, they display a considerably high mtDNA heterogeneity among them, namely in the frequencies of the U6, East African, and sub-Saharan haplogroups. In this study, we describe and compare the female gene pools of two neighboring Tunisian populations, Kesra (Berber) and Zriba (non-Berber), which have contrasting historical backgrounds. Both populations presented lower diversity values than those observed for other North African populations, and they were the only populations not showing significant negative Fu's F_s values. Kesra displayed a much higher proportion of typical sub-Saharan haplotypes (49%, including 4.2% of M1 haplogroup) than Zriba (8%). With respect to U6 sequences, frequencies were low (2% in Kesra and 8% in Zriba), and all belonged to the subhaplogroup U6a. An analysis of these data in the context of North Africa reveals that the emerging picture is complex, because Zriba would match the profile of a Berber Moroccan population, whereas Kesra, which shows twice the frequency of sub-Saharan lineages normally observed in northern coastal populations, would match a western Saharan population except for the low U6 frequency. The North African patchy mtDNA landscape has no parallel in other regions of the world and increasing the number of sampled populations has not been accompanied by any substantial increase in our understanding of its phylogeography. Available data up to now rely on sampling small, scattered populations, although they are carefully characterized in terms of their ethnic, linguistic, and historical backgrounds. It is therefore doubtful that this picture truly represents the complex historical demography of the region rather than being just the result of the type of samplings performed so far.

Both genetic data (e.g., Plaza et al. 2003) and historical data (El Heni 1991) support the distinction of North African populations from the rest of the continent,

[1]Laboratory of Molecular Genetics, Immunology, and Biotechnology, Faculty of Sciences of Tunis, University of Tunis, El Manar II 1060, Tunisia.

[2]Instituto de Patologia e Imunologia Molecular (IPATIMUP), Universidade do Porto, R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal.

[3]Faculdade de Ciências, Universidade do Porto, Pr. Gomes Teixeira, 4099-002 Porto, Portugal.

Human Biology, February 2005, v. 77, no. 1, pp. 61–70.

KEY WORDS: mtDNA VARIATION, TUNISIA, BERBERS, ARABS.

connecting them to Eurasians rather than to sub-Saharans. Available evidence points to a common Near Eastern source population settling both the north and the south coasts of the Mediterranean. On the mitochondrial level, two markers were advanced for tracing these parallel migrations: the sister clades U5, which migrated toward the west along the northern Mediterranean coast, and U6, which followed a similar direction but along the southern Mediterranean coast (Macaulay et al. 1999). The U6 mtDNA haplogroup has since been identified as characteristic of Berber populations, which are considered the North African ancestral population (Brett and Fentress 1997). However, the frequencies of this haplogroup among North African populations are highly heterogeneous (Côrte-Real et al. 1996; Rando et al. 1998; Brakez et al. 2001), possibly reflecting a differential population influence initiated from the 7th century onward with the rise of Islam. The Islamization of North Africa, which began in the east, was very fast: In less than a century the Berbers were the main component of the Islamic troops that invaded the Iberian peninsula, continuing the Islamic expansion to the northern Mediterranean coast [this peninsula remains the only region in Europe where U6 lineages have been observed in significant proportions (Côrte-Real et al. 1996; Salas et al. 1998; Pereira et al. 2000)].

Therefore, on the mtDNA level the picture that is emerging for North Africa is a complex one. Frequencies for the U6 haplogroup in Berber- and Arab-speaking neighboring populations in Morocco are not dissimilar (Rando et al. 1998) and are generally low (less than 10%), with maximum values of 20% in Mauritanians (Rando et al. 1998; from different ethnic groups) and 28% in Mozabite Berbers from Algeria (Côrte-Real et al. 1996). A slight west-east cline is nevertheless discernible, but we must remember that western North Africa is much better studied than eastern North Africa. There also has been sub-Saharan gene flow from sub-Saharan Africa (haplogroups L1, L2, and L3), which follows a south to north cline in North Africa (Plaza et al. 2003) and an east to west cline for the Ethiopian haplogroup M1 (Quintana-Murci et al. 1999).

Until now, research has focused on a broad North African scale, which has been further extended to comparison with the Iberian peninsula (Rando et al. 1998; Plaza et al. 2003). It is our aim in this study to describe and compare the female gene pools of two neighboring (40 km apart) but quite distinct central Tunisian populations—Kesra (Berber) and Zriba (non-Berber)—and to analyze them in the context of the mtDNA North African landscape. The two study populations have not interbred, at least not in recent times (El Heni 1991), and they have contrasting historical backgrounds. Kesra had continuous population settlement, starting in Berber times, and escaped the Bedouin conquest by Beni Hilel in the 11th century, which culminated in the Arabization of a large fraction of the Berber population (pushed to the mountains); Zriba was founded comparatively more recently with the settlement of a few families from Morocco about 400 years ago (El Heni 1991).

Today in Tunisia, the two branches of the Afro-Asiatic linguistic family can be identified but in largely disproportionate amounts. The Semitic branch is

represented by the Arabic language with many dialects, and the Berber branch is represented by the Tachelhit language, which is spoken by 2–3% of the Tunisian population (located in isolated groups in southern villages, such as Jerba island, Matmata, Sned, Douiret, Chenenni, and Zraoua). In northern Tunisia, Berbers have lost their language and speak Arabic (Grimes and Grimes 2004).

Materials and Methods

Population Samples. We analyzed two populations from central Tunisia: 47 unrelated individuals from the village of Kesra (Berbers) and 50 unrelated individuals from Zriba (Arabs). Only 40 km separate the two populations. A blood sample from each individual was collected in a test tube containing EDTA and was stored at −20°C.

DNA Extraction, Amplification, and Sequencing. DNA was extracted using a salting-out protocol (Miller et al. 1988). The conditions for D-loop HVRI PCR amplification were 30 cycles at 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min; the primers we used were L15997 (5′-CACCATTAGCACCCAAAGC-3′) and H16401 (5′-TGATTTCCGGAGGATGGTG-3′). The amplification products were purified with the QIAquick PCR purification kit (Qiagen). The sequencing reaction was carried out using the Big Dye Terminator cycle-sequencing ready-reaction kit (AB Applied Biosystems) and was performed separately on each strand with the primers L15997 and H16401, by using the sequencing profile of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min for 25 cycles. The products of the sequence reaction were precipitated in absolute ethanol and run in ABI 377 and ABI 3100 (AB Applied Biosystems) automatic sequencers.

Statistical Analysis. The nucleotide positions considered for analyses were between np 16024 and np 16383, according to Anderson et al. (1981). Sequences were aligned using the PHYLIP program, version 3.2 (Felsenstein 1989), and an input file for Arlequin (Schneider et al. 2000) was established for the calculation of gene diversity, mean pairwise differences, and neutrality tests. The distance between populations was estimated using the intermatch-mismatch models using the equation

$$D = d_{ij} - \frac{(d_{ii} + d_{jj})}{2} \tag{1}$$

(Nei 1987), where d_{ij} is the distance between populations i and j and d_{ii} and d_{jj} are the intrapopulation distances.

Control region sequences were assigned to haplogroups according to the nomenclature of Richards et al. (1998), Rando et al. (1998), and Macaulay et al.

(1999). Principal coordinates analysis of intermatch-mismatch genetic distances was performed using the SAS Statistical Analysis System, version 6.12.

Results

Several measures of sequence diversity (Table 1) show that the Kesra population presents a higher diversity than the Zriba population, although the value is still lower than the values observed for other North African populations. The only exception is the mean number of pairwise differences for Kesra, which is as high as the values observed for other North African populations. Both Kesra and Zriba are the only North African populations not showing significant negative Fu's F_s values, suggesting a substantial population substructure as well as considerable drift, particularly in Zriba, as reflected by the relatively high frequencies of some haplotypes (16248; 16311-16391; $16182^{A/C}$-$16183^{A/C}$-16189-16234-16324).

Remarkably, the two populations shared only three haplotypes (see Appendix): the CRS sequence, a haplotype with the transition at position 16248, and a haplotype with transitions at positions 16209-16218-16223-16292-16311.

The haplogroup assignment clarified the reason for the higher diversity of the Kesra population (Table 2), with both sets of typical sub-Saharan and Eurasian haplotypes each amounting to 49%, which raises the mean number of pairwise differences. In contrast, in Zriba sub-Saharan sequences made up only a small proportion (8%) of the haplotypes. Of the sub-Saharan set of sequences, haplotypes belonging to the M1 haplogroup, which is highly frequent in Ethiopia, were observed only in Kesra, with a frequency of 4.2%.

With respect to the U6 haplogroup, frequencies were low, amounting to 2% in Kesra and 8% in Zriba, and all belonged to the subhaplogroup U6a.

In the context of North Africa (Figure 1), these haplogroup data present a picture that is quite complex. Zriba, the Tunisian non-Berber population, would match the profile of a Berber Moroccan population. In contrast, Kesra, considered Berber, has twice the sub-Saharan haplogroup frequency that would be expected in such a northern population and would match a western Saharan population except for the low U6 frequency.

It is not surprising therefore that an interpretation of the results of a principal coordinates plot for North African population genetic distances is difficult to achieve (Figure 2). Although the first two axes explain 76.52% of the total variance, the genetic data and the ethnic or geographic information show no association, reinforcing the notion of a complex history of North African populations, as suggested in previous studies (Brakez et al. 2001; Krings et al. 1999; Plaza et al. 2003; Rando et al. 1998).

Discussion

The expectation that the increasing number of published studies on the mtDNA phylogeography of North Africa would bring clarity to the complex

Table 1. HVRI (from np 16024 to np 16365) Diversity and Neutrality Measures in North African Populations

North African Population	*Sample Size*	*Haplotype*[a]	*Segregating Sites*[b]	*Gene Diversity*[c]	*Mean Pairwise Differences*	*Tajima's D*	*Fu's F_s*	*Reference*
Kesra	47	20 (42.6)	43 (11.9)	0.932 ± 0.021	6.265	−1.23	−3.67	This study
Zriba	50	16 (32.0)	31 (8.6)	0.904 ± 0.022	3.949	−1.43	−3.22	This study
Algeria	86	31 (36.0)	37 (10.3)	0.945 ± 0.010	4.822	−1.09	−13.2[e]	Côrte-Real et al. (1996)
Bedouin	29	27 (93.1)	51 (14.2)	0.995 ± 0.011	7.512	−1.58[d]	−20.6[e]	Di Rienzo and Wilson (1991)
Egypt	68	59 (86.8)	71 (19.7)	0.993 ± 0.005	7.075	−1.77[d]	−25.0[e]	Krings et al. (1999)
Mauritania	30	23 (76.7)	30 (8.3)	0.975 ± 0.017	6.025	−0.74	−13.0[e]	Rando et al. (1998)
Morocco Berber	60	38 (63.3)	48 (13.3)	0.963 ± 0.015	4.594	−1.86[d]	−25.7[e]	Pinto et al. (1996)
Morocco non-Berber	32	29 (90.6)	46 (12.8)	0.988 ± 0.014	6.026	−1.73[d]	−24.8[e]	Rando et al. (1998)
Souss	50	34 (68.0)	38 (10.6)	0.961 ± 0.018	4.604	−1.55[d]	−25.6[e]	Brakez et al. (2001)
Tuareg	23	21 (91.3)	40 (11.1)	0.992 ± 0.015	6.838	−1.43	−13.7[e]	Watson et al. (1997)
Western Saharan	25	20 (80.0)	31 (8.6)	0.973 ± 0.022	5.340	−1.31	−11.9[e]	Rando et al. (1998)

a. Number of distinct haplotypes in sample (percentage of sample size).
b. Number of sites variable in sample (percentage of all sites).
c. Average heterozygosity ± standard error.
d. $0.01 < p \leq 0.05$.
e. $p \leq 0.001$.

Table 2. Haplogroup Distributions (%) in Kesra and Zriba

Population	*Haplogroup*	*Kesra*	*Zriba*
Eurasians	H + HV	29.8	60.0
	K	–	2.0
	V	6.4	–
	U (except U6)	4.3	16.0
	T	4.3	6.0
	J	2.1	–
	X	2.1	8.0
North Africans	U6a	10.6	–
Sub-Saharans	L1	17.0	4.0
	L2	17.0	4.0
	L3	4.3	–
	M1		

population history of this region is not being met. This is most likely because most previous genetic studies in this region have considered only small dispersed populations, despite careful selection of sample groups from cultural, linguistic, and historical viewpoints.

This applies obviously to this study, in which a high level of intrapopulation structure and a low exchange of sequences between neighboring populations of different ethnicity are evidenced. The non-Berber population of Zriba had a surprisingly high frequency for the U6 haplogroup, but the absence of M1 sequences and the low frequency of L sequences could support its affiliation with

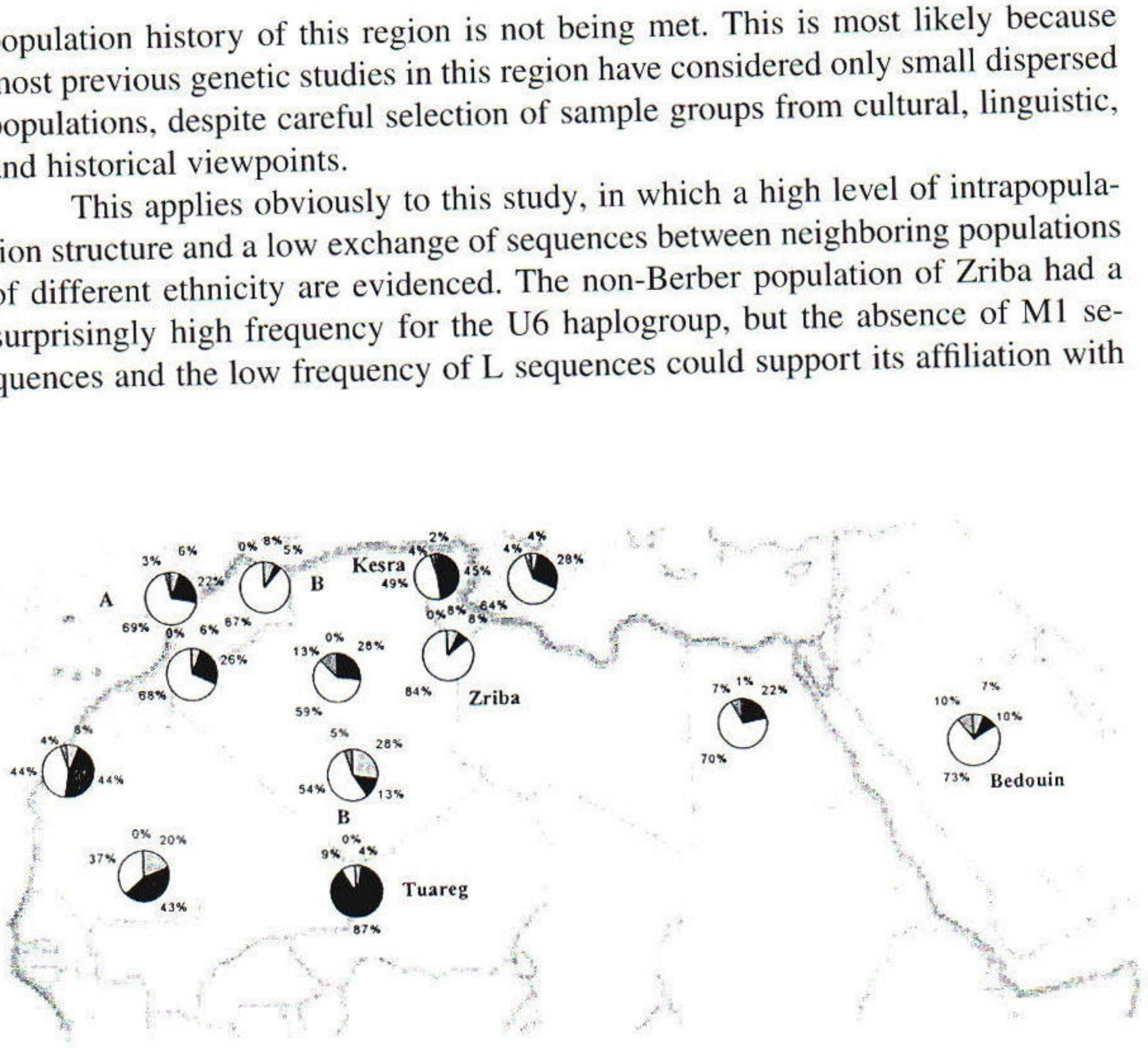

Figure 1. Frequencies of North African (U6 haplogroup) (light gray), sub-Saharan (haplogroups L1, L2, and L3) (black), Ethiopian (M1) (dark gray), and Eurasian (white) lineages in North African populations. B, Berber; A, Arab.

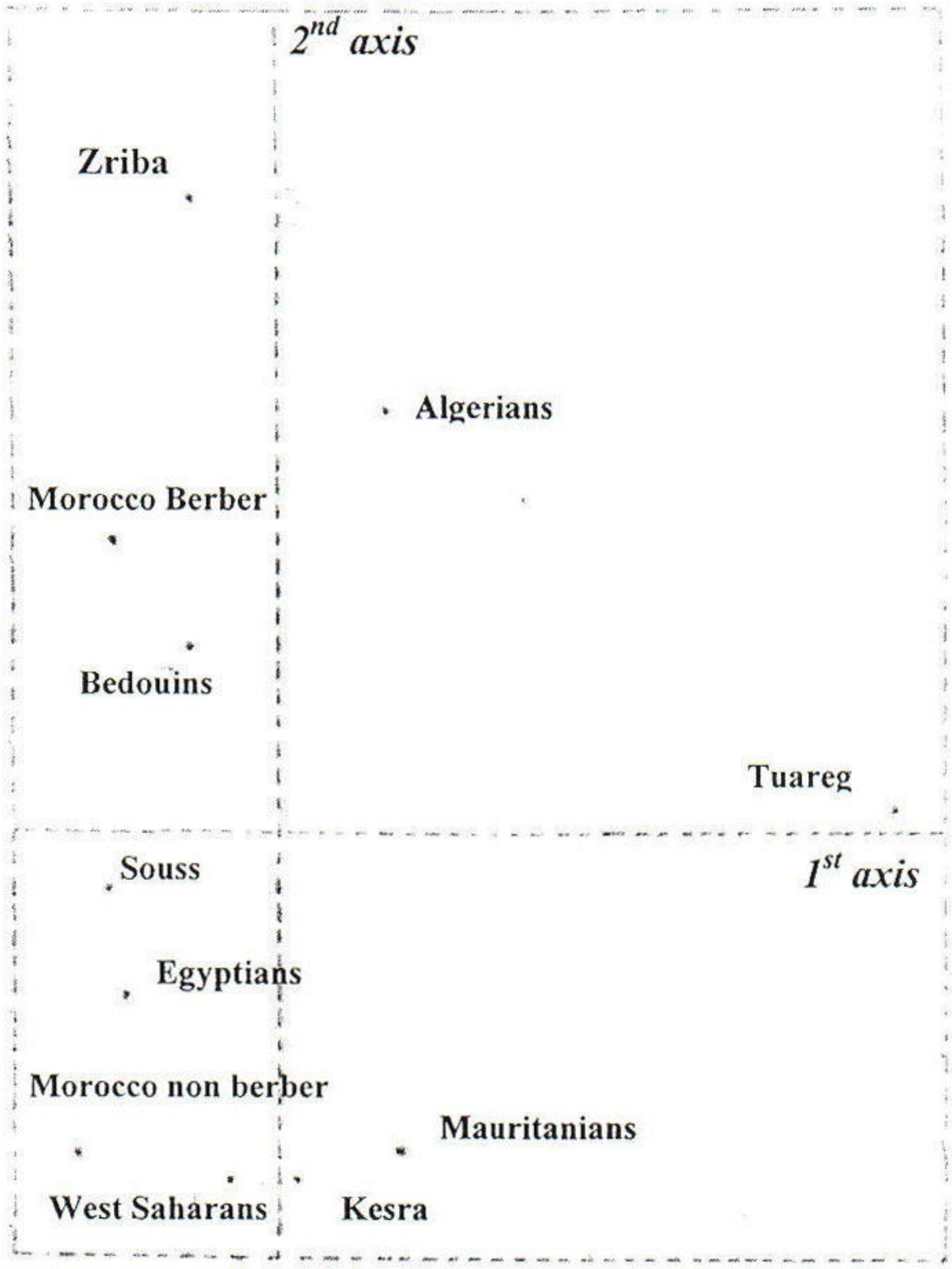

Figure 2. First and second principal coordinates estimated from the intermatch-mismatch distances among North African populations. The first two principal coordinates explain 76.66% of the variation.

Morocco, although a Berber background would be more likely than an Arab one. Kesra, although Berber, showed a low frequency of the U6 haplogroup and an unexpectedly high frequency of L sequences for a Mediterranean North African population. A comparison of Arab and non-Arab populations in the Middle East (Richards et al. 2003) showed that the first populations presented higher frequencies of the L haplogroup (10–35%), pointing to high gene flow in female sub-Saharan lineages to Arab populations as a result of slave trade in the last 2,500 years. In North Africa the scenario is quite different, evidencing quite variable proportions of sub-Saharan lineages among Arab and Berber communities. The estimates for the proportion of the L haplogroup in North African populations

vary from 5% to 45% (and 87% in the nomadic Tuaregs), but again a complex picture emerges if we take into account ethnicity and geography.

At present no other area of the world has been described as having a patchy mtDNA landscape like the one found in North Africa. Does this imply that substructure in this region is deeper than in other regions? And if so, what is the cause? Present evidence does not match any coherent ethnic or geographic explanatory framework, and probably new sampling strategies will need to be considered.

Acknowledgments This study was supported by Fundação para a Ciência e a Tecnologia (Eixo 2, Medida 2.3 do POCTI, QCA III) and the Tunisian Ministry of Higher Education, Scientific Research, and Technology through project POCTI/ANT/45139/2002. L. Pereira received a research grant (SFRH/BPD/7121/2001) from Fundação para a Ciência e a Tecnologia. The Instituto de Patologia e Imunologia Molecular of the Universidade do Porto (IPATIMUP) is financed by Programa Operacional Ciência, Tecnologia e Inovação (POCTI), Quadro Comunitário de Apoio III. We are grateful to the Tunisian Ministry of Health and the staff of the Kesra and Siliana hospitals for their help in collecting blood samples and to two anonymous referees for considerable improvements to the manuscript.

Received 12 February 2004; revision received 27 April 2004.

Literature Cited

Anderson, S., A. T. Bankier, B. G. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.

Brakez, Z., E. Bosch, H. Izaabel et al. 2001. Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann. Hum. Biol.* 28:295–307.

Brett, M., and E. Fentress. 1997. *The Berbers.* Oxford: Blackwell.

Côrte-Real, H. B., V. A. Macaulay, M. B. Richards et al. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.* 60:331–350.

Di Rienzo, A., and A. C. Wilson. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 88:1,597–1,601.

El Heni, A. 1991. *Safahat min tarikh siliana* [Pages from Siliana History]. Tunis, Tunisia: Dar Cheima linacher wa taouzia, 7–9.

Felsenstein, J. 1989. PHYLIP: Phylogeny inference package (Version 3.2). *Cladistics* 5:164–166.

Grimes, B. F., and J. E. Grimes, eds. 2004. *Ethnologue: Languages of the World,* 14th ed. Available at http://www.ethnologue.com

Krings, M., A. E. Salem, K. Bauer et al. 1999. mtDNA analysis of Nile River valley populations: A genetic corridor or a barrier to migration? *Am. J. Hum. Genet.* 64:1,166–1,176.

Macaulay, V., M. Richards, E. Hickey et al. 1999. The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* 64:232–249.

Miller, S. A., D. D. Dykes, and H. F. Polesky. 1988. A simple salting-out procedure for extracting DNA from nucleated cells. *Nucleic Acids Res.* 16:1,215.

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

Pereira, L., M. J. Prata, and A. Amorim. 2000. Diversity of mtDNA lineages in Portugal: Not a genetic edge of European variation. *Ann. Hum. Genet.* 64:491–506.

Pinto, F., A. M. Gonzalez, M. Hernandez et al. 1996. Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann. Hum. Genet.* 60:321–330.

Plaza, S., F. Calafell, A. Helal et al. 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., O. Semino, H.-J. Bandelt et al. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.* 23:437–441.

Rando, J. C., F. Pinto, A. M. Gonzalez et al. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, Near Eastern, and sub-Saharan populations. *Ann. Hum. Genet.* 62:531–550.

Richards, M. B., V. A. Macaulay, H.-J. Bandelt et al. 1998. Phylogeography of mitochondrial DNA in Western Europe. *Ann. Hum. Genet.* 62:241–260.

Richards, M., C. Rengo, F. Cruciani et al. 2003. Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am. J. Hum. Genet.* 72:1,058–1,064.

Salas, A., D. Comas, M. V. Lareu et al. 1998. mtDNA analysis of the Galician population: A genetic edge of European variation. *Eur. J. Hum. Genet.* 6:365–375.

Schneider, S., D. Roessli, and L. Excoffier. 2000. *Arlequin ver. 2.000: A Software for Population Genetics Data Analysis.* Geneva: Genetics and Biometry Laboratory, University of Geneva.

Watson, E., P. Forster, M. Richards et al. 1997. Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* 61:691–704.

Appendix 1. Control Region Sequences in Kesra and Zriba[a]

Kesra	*Zriba*	*HVRI*	*Haplogroup*
10	11	CRS	H
	1	148	H
	2	$183^{A/C}$ 189	H
3	5	248	H
	3	263 288 362	H
1		304	H
	2	262 311 391	HV
	6	311 391	HV
1		069 126 145 222 256 261	J1b
	1	129 224 311	K
1		129 148 168 172 187 $188^{C/G}$ 189 223 230 278 293 311	L1a
4		126 187 189 223 264 270 278 293 311	L1b
4		092 209 214 223 278 294 301 354 390	L2a
	2	129 189 192 223 278 287 294 309 390	L2a
1		$183^{A/C}$ 189 223 278 294 362	L2a
2		189 223 278 294 362	L2a
1		209 214 223 278 294 301 354 390	L2a
	1	179 215 223 $256^{C/A}$ 284 311	L3*
3	1	209 218 223 292 311	L3*
2		223 286 311 320	L3e2
3		223 320 399	L3e2
1		129 $183^{A/C}$ 189 223 249 291 311	M1

Appendix 1. (Continued)

Kesra	*Zriba*	*HVRI*	*Haplogroup*
1		129 183$^{A/C}$ **189** 223 249 311	M1
	3	126 183^{delA} 294 296 304	T*
2		126 163 186 189 261 294	T1
1		162 343	U3
1		093 189 270	U5b
	1	172 180 219 278	U6a
1		172 183$^{A/C}$ 189 219 239 278	U6a1
	1	172 183$^{A/C}$ 189 219 278	U6a1
	2	172 183$^{A/C}$ 189 219 278 362	U6a1
	8	182$^{A/C}$ 183$^{A/C}$ 189 234 324	U8b
3		298	V
1		182$^{A/C}$ 183$^{A/C}$ 189 223 278	X
47	50		

a. Variant positions from the CRS are shown between np 16024 and np 16391 in HVRI (minus 16000). The substitution indicated in boldface was observed to be heteroplasmic. Substitutions are transitions unless the base change or a deletion is explicitly indicated.