

MtDNA Profile of West Africa Guineans: Towards a Better Understanding of the Senegambia Region

Alexandra Rosa^{1,2}, António Brehm^{2*}, Toomas Kivisild¹, Ene Metspalu¹ and Richard Villems¹

¹*Department of Evolutionary Biology, Estonian Biocenter, Tartu University, Riia 23, 51010 Tartu, Estonia*

²*Human Genetics Laboratory, Center of Macaronesian Studies, University of Madeira, Campus of Penteadá, 9000-390 Funchal, Portugal*

Summary

The matrilineal genetic composition of 372 samples from the Republic of Guiné-Bissau (West African coast) was studied using RFLPs and partial sequencing of the mtDNA control and coding region. The majority of the mtDNA lineages of Guineans (94%) belong to West African specific sub-clusters of L0-L3 haplogroups. A new L3 sub-cluster (L3h) that is found in both eastern and western Africa is present at moderately low frequencies in Guinean populations. A non-random distribution of haplogroups U5 in the Fula group, the U6 among the “Brame” linguistic family and M1 in the Balanta-Djola group, suggests a correlation between the genetic and linguistic affiliation of Guinean populations. The presence of M1 in Balanta populations supports the earlier suggestion of their Sudanese origin. Haplogroups U5 and U6, on the other hand, were found to be restricted to populations that are thought to represent the descendants of a southern expansion of Berbers. Particular haplotypes, found almost exclusively in East-African populations, were found in some ethnic groups with an oral tradition claiming Sudanese origin.

Introduction

Unveiling the history of human settlement in the West Coast of Africa is a complex task. It is the result of a continuous complex network of migrations, invasions and admixture of peoples from different origins. Fossil evidence suggests a modern human presence in NW Africa around 40000 years before present (YBP) (Alimen, 1987). A pre-Neolithic Capsian culture evolved later locally or through a diffusion from the Near East (Camps-Faber, 1989). Around 9000 YBP, when the Sahara went through a period of maximum humidity (Aumassip *et al.* 1988), several Neolithic cultures flourished in the area, bringing together people of sub-Saharan and North African origin (Dutour *et al.* 1988). The domestication and spread of several African-specific plants probably started in western Sahel after 4000 YBP. The first phase of largely east and southward oriented Bantu migrations, originating from the cen-

tral Gulf of Guinea region, is a likely outcome of these cultural developments (Fage, 1995).

The Ghana Empire, between Niger and Senegal, is the oldest known occidental African Kingdom (Fage, 1995) which was followed in the 14th–16th centuries by other empires (Mali, Songhai). The admixture of Berbers with native populations of this area dates back at least to the 9th century A.D., after the arrival of pastoral Peuls or Fulbe (here designated as Fula). In 1086 Ommiádes conquered North-Western Africa and pushed the populations from South Morocco and Mauritania to the Senegal region (Moreira, 1964). When the Europeans arrived in Senegambia in the 15th century they met most of the presently known ethnic groups settled in the region (Teixeira da Mota, 1954). The Fula arrived again two centuries later, coming from the Futa Toro and Sahel regions, dominating the whole area. The Mandinga (Mandenka) were the last to arrive in this region (Carreira & Quintino, 1964).

Present day Guinean ethnic groups are disseminated all over the territory. The Balanta are the biggest group, and in the first quarter of the 20th century spread over

*Corresponding author: António Brehm, phone +351291705383, fax +351291705393, e-mail: brehm@uma.pt

territories occupied earlier by other ethnic groups. The origin of the Balantas is uncertain. Some see language affinities with the Sudanese from whom they could have separated 2000 years ago with the first spread of Kushites migrations (Quintino, 1964). According to Stuhlmann (1910), the group derives from a Bantu branch, which separated in the Pleistocene near the Nile, following camite invasions. The Bijagós inhabit the Archipelago of the same name and some scholars see strong cultural resemblances to Egyptians (Quintino, 1964), but others relate them to the Senegalese Djola. The latter are a rather heterogeneous group, and include the Beafada which have an oral tradition of coming from Mali (Lopes, 1999). A mass arrival of Fula took place in the beginning of the 19th century. The origin of this ethnic group is unknown, but tradition relates them to Hiksos and Nubians. They show the typical phonetic “glottal catch” which characterizes the whole group.

Here we analyze the mtDNA lineages present in the major ethnic groups of Senegambia, covering a broad number of recognized groups underrepresented in previous studies (Graven *et al.* 1995; Watson *et al.* 1997; Rando *et al.* 1998), and compare them within the broader context of African mtDNA variability (Graven *et al.* 1995; Watson *et al.* 1997; Rando *et al.* 1998, 1999; Krings *et al.* 1999; Chen *et al.* 2000; Pereira *et al.* 2001; Brehm *et al.* 2002; Salas *et al.* 2002). Because mtDNA haplogroups show distinct geographic patterns in Africa, their frequency and diversity patterns in West Africa can be informative with respect to the origin of the different ethnic groups from Guiné-Bissau. The presence of Y-chromosomes of Eurasian affiliation among populations from Cameroon at a high frequency, as reported recently (Cruciani *et al.* 2002), raises the intriguing question of back migrations from Eurasia to Africa, here supported by the presence of particular Eurasian mtDNA lineages among Guineans.

Material and Methods

Sampling

A total of 372 blood samples were collected from unrelated Guinean males whose maternal ancestors were known to belong exclusively to a specific ethnic group. The samples were collected either in military camps

with the permission of the Guiné-Bissau Chairman of the Joint Chiefs of Staff, or in the villages around the country with the help of the Ministry of Health. Every participant gave his consent in an individual interview after a detailed explanation of the project. Sample sizes and origins (along with additional information) are specified in Table 1 and 2. Due to the complex history involving the major ethnic groups in Guiné-Bissau, they do not all follow a clear present-day settlement pattern (see Figure 1).

Populations of low sample size were pooled according to their linguistic affinities. The linguistic clustering presented in Table 1 is based on anthropological or linguistic classifications following Almeida (1939), Barros (1947), Carreira (1962, 1983), Almada (1964), Carreira & Quintino (1964), Hair (1967), Quintino (1967, 1969), Diallo (1972) and Lopes (1999). Some groups were left unpooled: the Balanta, for whom a Sudanese origin has been suggested, and the Bijagós because of their particular geographical location.

HVS-I and HVS-II Sequencing

The leukocyte fraction of whole blood was used for DNA extraction by standard methods and the mtDNA hypervariable segment I (HVS-I) of the control region was amplified and sequenced. Sequencing products were separated on a MegaBACE 1000 automatic sequencer, following the manufacturer's specifications and aligned using Wisconsin Package GCG Version 10.0. All sequences were read between nucleotide positions (nps) 16024 and 16400. Additional information regarding polymorphic sites 185, 186, 189, 195, 236, 297 and 322 in HVS-II was obtained by directly sequencing all samples that could not be unambiguously classified on the basis of HVS-I information alone.

RFLP Testing

In case of ambiguity in defining mtDNA haplogroups on the basis of the HVS-I haplotype, additional data was gathered from restriction fragment length polymorphisms (RFLPs) of diagnostic sites. All restriction digests were made according to the manufacturer's instructions (Fermentas and New England BioLabs). The following polymorphic restriction sites were screened: 322HaeIII,

Table 1 Population data of the Guinean samples ethnic distribution

Code	Ethnic group	Language group WA	Religion	Closest language group	Synonyms
BLE	Balanta ^a	Bak-Balanta-Ganja	A,M,C	Tenda	Ballante, Balant
BDA	Baiote	Bak-Diola-Bayot	A	Diola	Bayotte
BAB	Banhu	Eastern Senegal-Banyun	A,M	Tenda	Bainouk, Banyuk, Elomay
BIF	Beafada	Easter Senegal-Tenda	M	Badyara	Biafada, Bidyola, Biafar
BJG	Bijagó	Bijagó	A		Bidyogo, Bijougot
BRA	Brame				
CCJ	Cassanga	Eastern Senegal-Nun	A	Banhu-Felupe	Kasanga, I-Hadja
EJA	Djola ^b	Bak-Diola-HerEjamat	A,C	Diola-Wolof	Fulup, Floop, Ejamat, Ediamat
FUL	Fula	Fulani-West Central	M	Fula-Wolof	Fulbe, Futa Jallon
FUF	Futa-Fula	Fulani-West Central	M	Fula-Wolof	Fulbe, Futa Jallon
FUC	Fula-Preto	Fulani-Western	M	Fula-Wolof	Peul, Peulh
FUC	Fula-Forro	Fulani-Western	M	Fula-Wolof	Peul, Peulh
FUT	Fula-Toranca	Fulani	M	Fula-Wolof	Peul, Peulh
JAD	Jancauca	Mandenkan	M	Mandinka	Jahanque, Jahanka, Diakanke
LAN	Landoma				
MAN	Mancanha	Bak-Manjaku-Papel	A,C	Manjaku-Papel	Mankanya, Mankanha
MNK	Mandinga	Mandekan	M	Kalenke, Jahanka	Mandingue, Mandenka
MFV	Manjaco	Bak, Manjaku-Papel	A,C,M	Mancanha, Papel	Mandyak, Manjiak
MSW	Mansonca	Sua	M		Kunante, Mansoanka
NAJ	Nalú	Mbulungish-Nalu	A,M	Susu	Nalou
SUD	Sussu	Susu-Yalunka	M,A,C	Yalunka	Susu, Sose, Soso
PBO	Papel	Bak, Manjaku-Papel	A,C	Mankanya, Mandyak	Pepel, Oium

A-Animist, M-Muslim, C-Christian; Population codes and language groups follow terminology from www.sil.org/ethnologue/; ^aincludes the so-called Balanta-Mané (Balanta islamized by Mandinga); ^bincludes Felupes;

1715DdeI, 2349MboI, 2758RsaI, 3592HpaI, 3693MboI, 4157AluI, 4685AluI, 5584AluI, 5656NheI, 7055AluI, 8616MboI, 10084TaqI, 10321AluI, 10394DdeI, 10397AluI, 10806HinfI, 11439MboI, 11641HaeIII, 12308HinfI, 13803HaeIII, 13957HaeIII, 14766MseI and 14868MboI. The following coding region sites were ascertained by sequencing: 2758, 4218, 12618, 13105 and 14182. Primers and PCR conditions used in all analyses are available as Complementary Material at www.ahg.com.

Haplogroup characterization

The HVS-I sequence types were classified following the nomenclature of African and European mtDNA haplogroups (Quintana-Murci *et al.* 1999; Macaulay *et al.* 1999; Rando *et al.* 1999; Alves-Silva *et al.* 2000; Chen *et al.* 2000; Richards *et al.* 2000; Richards & Macaulay 2001; Bandelt *et al.* 2001; Torroni *et al.* 1997, 2001; Mishmar *et al.* 2003; Salas *et al.* 2002). Here, and in what follows, the nucleotide position (np) number relative to

the revised CRS (Anderson *et al.* 1981; Andrews *et al.* 1999) is used to designate haplotype-defining mutations. Character state change is specified only for transversions and insertions/deletions. Based on the previous knowledge of African complete sequences paraphyletic clade L1 is split into two monophyletic units L0, capturing previously defined L1a and L1d lineages, and L1 clade that includes L1b and L1c clades (Mishmar *et al.* 2003). The sub-clades of L0a (*pro* L1a) and L1b are defined as in Salas *et al.* (2002).

Haplogroup L2 is divided into L2a (characterized by 16294 and 13803), L2b (16114A, 16129, 16213 and 4158), L2c (322 and 13958), and L2d (16399 and 3693) sub-clades. Mutations 16278, 16362 and 10086 characterize haplogroup L3b; haplogroup L3d is defined by 8618 and shares with L3b a transition at np 13105. According to Bandelt *et al.* (2001) L3e (defined by 2352) is subdivided into L3e1 (16327), L3e2 (16320), L3e3 (16265T) and L3e4 (16264 and 5584) clades. L3e2 is further subdivided into L3e2* (14869) and L3e2b (16172 and 16189). As in Salas *et al.* (2002), L3f captures all L3*

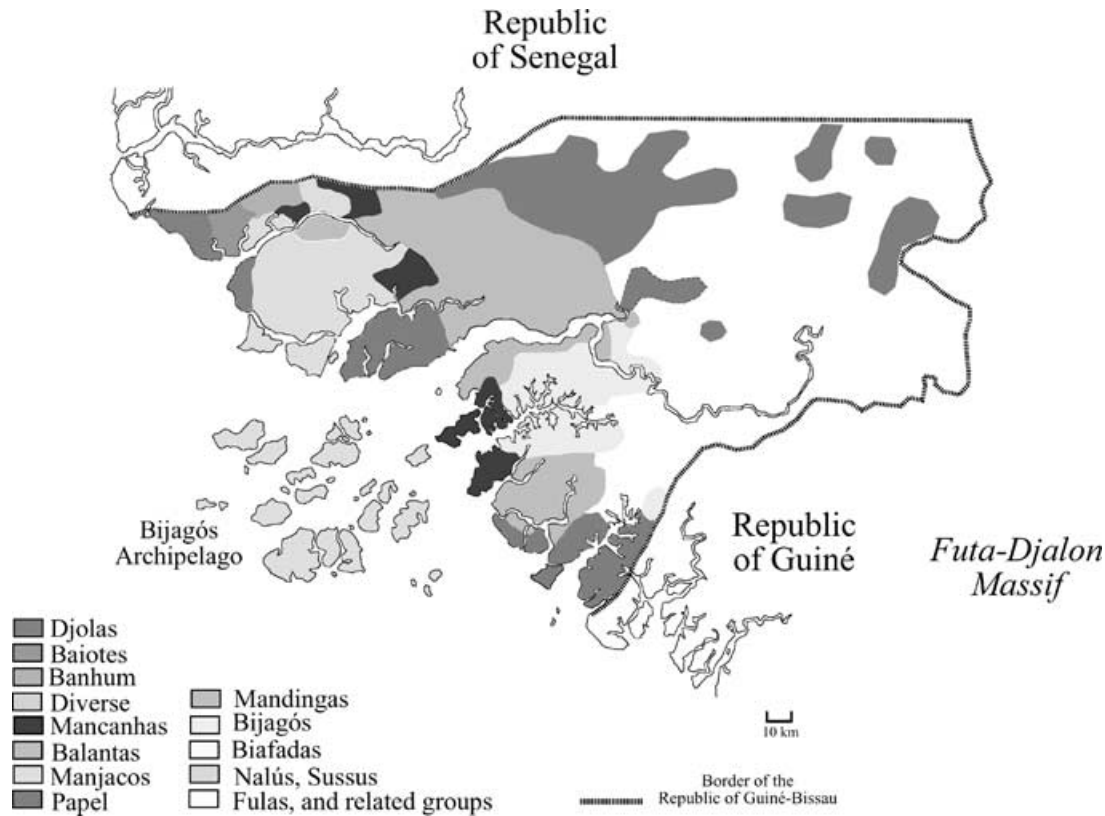


Figure 1 Geographic distribution of ethnic groups in Guiné-Bissau. The boundaries may not correspond entirely to the precise distribution of the groups involved since overlapping areas do exist.

lineages with a mutation at 16209. L3f1 is further defined by a T at np 16292 (and 14766). Here we further define a new sub-cluster, L3h characterized by a loss of the *DdeI* site at np 1715 (mutation at np 1719) and the HVS I motif 16129, 16256A and 16362. Following Finnilä *et al.* (2000) U5b is characterized by 5656 and 12618 over 14182. Haplogroup U6 (Rando *et al.* 1998) is defined by 16172 and 16219. Haplogroup M1 is characterized by 16129, 16189, 16249 and 10400 mutations (Quintana-Murci *et al.* 1999).

Genetic Analysis and Population Comparisons

Median networks of HVS-I haplotypes (Bandelt *et al.* 1995, 2000) were drawn for each haplogroup separately, using the Network 3.1 program (Arne Röhl, www.fluxus-engineering.com/sharenet.htm). Haplogroup frequencies, molecular diversity indexes (F_{ST}) and genetic diversity ($H - Nei$, 1987) for haplotypes and haplogroups and analysis of molecular variance (AMOVA) were calculated using Arlequin

v2.0 (Schneider *et al.* 2000). Comparisons between populations were assessed by subjecting the (relative) frequency vectors of the haplogroups to a principal component analysis (PCA).

A local database with more than 19000 individuals taken from literature and our unpublished data from worldwide populations was employed to search for exact matches of Guiné-Bissau haplotypes, ignoring length variation in the C stretch of the HVS-I.

Coalescence times were estimated by means of the ρ statistic, assuming that a transition within 16090–16365 corresponds to 20180 years (Forster *et al.* 1996).

Results and Discussion

Haplogroup Profiles

The 372 Guinean samples clustered to 192 different haplotypes of all major West African mtDNA haplogroups (for the complete list see Complementary Material). Three predominant haplotypes (GB4, GB85

Table 2 Haplogroup relative frequencies and diversity index (H) in Guiné-Bissau ethnic groups and several other African populations. Superscripts (a-g) in Guinean ethnic groups refer to codes used in Figure 3 (PCA)

Haplogroup	Guiné-Bissau																						
	Mozambique (Mo) ^{1,2}	Ethiopia (Et) ³	Kenya/Sudan ^{4,i}	Egypt (Eg) ⁵	Nile Valley (Nv) ⁶	Khung/ ^{7,8}	Verde (CV) ⁹	Senegal Mand (Sm) ¹⁰	Sen. mixed (Ss) ^{11,ii}	Niger/Nigeria (NN) ^{4,8,iii}	BAB/CCJ/BIF ^a	BDA/EJA/CCJ/BIF ^a	BJC ^b	BLE ^c	PBO/MFV/ MAN ^d	FUC/FUF/ FUL/FUT ^e	MNK/MSW/ LAN/ JAD/SUS ^f	NAJ ^g	Guiné Total	Morocco Arabs (Ma) ^{11,12}	Mor. Berbers (Mb) ^{11,12}	Algeria Berbers (Ab) ^{12,13}	Alg. Arabs (Aa) ^{11,12}
L0a	25	7	13	3	8	4	1	2	0	1	4	4	9	11	4	1	5	4	5	0	0	1	2
L1b	1	2	0	1	4	1	8	20	17	15	6	6	18	13	6	17	9	4	11	5	4	1	2
L1c	5	0	0	1	0	0	7	2	3	3	2	0	7	4	6	5	8	5	1	1	0	0	0
L1*	6	4	16	1	4	69	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
L2a	33	13	16	5	15	3	20	12	20	20	12	13	15	18	22	14	15	17	5	5	3	14	14
L2b	1	2	0	2	0	4	4	3	16	1	10	5	6	13	0	7	19	8	1	1	0	2	2
L2c	1	0	0	0	0	0	16	39	7	4	18	13	11	17	16	23	15	16	1	1	0	4	4
L2*	1	0	0	0	2	0	1	0	0	1	0	0	3	1	3	2	4	2	1	1	0	5	5
L3b	3	0	0	1	1	7	10	8	11	13	14	13	5	8	10	5	4	8	4	2	2	9	9
L3d	6	3	0	1	0	0	7	9	7	12	2	9	13	12	5	12	15	9	2	0	1	5	5
L3e	15	0	3	3	1	12	15	4	7	15	14	5	11	4	3	7	12	7	5	2	4	14	14
L3f	2	5	8	3	4	0	1	0	3	6	6	5	0	3	1	3	0	2	2	0	0	0	0
L3h	0	1	0	0	0	0	1	0	0	1	8	5	2	4	1	5	0	4	0	0	0	0	0
L3*	1	15	32	0	11	0	3	0	3	0	0	0	0	0	0	0	0	0	1	1	0	4	4
M1	0	17	5	7	8	0	0	0	0	0	4	5	3	0	0	0	0	1	4	3	4	0	0
U6	0	3	2	2	1	0	3	0	2	3	0	0	0	5	3	3	0	2	7	7	24	2	2
Eurasian	0	28	6	70	41	0	3	1	4	4	0	0	0	0	12	0	0	3	61	72	60	37	37
H	.795	.844	.828	.509	.781	.518	.884	.778	.881	.876	.905	.926	.909	.899	.876	.901	.902	.901	.610	.476	.573	.822	.822
sd	.011	.011	.022	.043	.020	.058	.007	.027	.012	.011	.016	.028	.011	.014	.016	.019	.027	.005	.029	.037	.037	.037	.037
N	416	270	88	192	255	93	292	110	121	137	50	22	62	77	77	58	26	372	350	268	149	55	55

¹Pereira et al. (2001); ²Salas et al. (2002); ³unpublished; ⁴Watson et al. (1997); ⁵unpublished; ⁶Krings et al. (1999); ⁷Chen et al. (2000); ⁸Vigilant et al. (1991); ⁹Brehm et al. (2002); ¹⁰Graven et al. (1995); ¹¹Rando et al. (1998); ^{12,13}unpublished (i – Turkana, Kikuyu and Somalia; ii – Senegalese, Wolof and Serer; iii – Songhai, Tuareg, Yoruba, Hausa, Fulbe and Kanuri). The so-called Eurasian haplogroup refers to all non-Ls, M1 or U6 sequences. N, total sample number; H, Nei's (1987) diversity index; sd, H standard deviation.

and GB117) captured 13% of the Guinean mtDNA variation, occurring at a frequency higher than 3% each. Most sequences (94%) could be classified as belonging to sub-Saharan African L0a1, L1b, L1c1, L2a, L2b, L2c, L2d1, L3b, L3d, L3e, L3f1 and L3h haplogroups and subhaplogroups. Unexpectedly for a West African population, 22 (5.9%) of the samples clustered to haplogroups M1 (1.1%), U5 (2.7%) and U6 (2.2%, Table 2; Graven *et al.* 1995; Watson *et al.* 1997; Rando *et al.* 1998; Salas *et al.* 2002). M1 and U6 are found in North and East Africa, Arabia, and the Middle East, whereas U5 has been sampled at appreciable frequencies only in Europe (Passarino *et al.* 1998; Quintana-Murci *et al.* 1999; Richards *et al.* 2000). The haplogroup profile for each ethnic group separately can be found in the Complementary Material.

L Lineages

Haplogroup L0 was represented in Guineans only by its daughter group L0a1 showing marginal frequencies ranging from 1% to 5% (Table 2), in contrast to its frequency in East African populations (e.g. 25% in Mozambique: Watson *et al.* 1997; Pereira *et al.* 2001; Salas *et al.* 2002). Interestingly, only the Balanta, a group claiming Sudanese origin, showed an increased frequency of this clade (11%). Haplogroup L0a has a Paleolithic time depth in East African populations (33,000 year old, Salas *et al.* 2002). The relatively young coalescent date of L0a1 in Guineans (6400 ± 2600 years, assuming a single founder) suggests that only a small subset of L0a reached Guinea during the Holocene. The founder haplotype of L0a in Guineans, GB4 (see Table 4 in Complementary Material), has an exact match in East Africa, the Middle East and in Cape Verde and Senegal Mandenka populations, indicating that its spread is not strictly restricted to Guineans. The lack of the L0a2 clade, associated with the 9bp deletion in CoII/tRNA^{Lys} intergenic region, and widespread in Bantu speaking populations all over Africa (Soodyall *et al.* 1996), suggests that L0a has at least two distinct phylogeographic patterns in Central and West Africa. We cannot discard the possibility of a Bantu migration to West Africa, as the founder group could have a distinct composition from those who participated in the southwards migration(s).

Haplogroup L1b is restricted mostly to West African populations (Graven *et al.* 1995; Watson *et al.* 1997; Salas *et al.* 2002) and is represented by two different branches in Guineans. Its major cluster (Figure 2) L1b1 is associated with a transition at np 16293 and includes a frequent sub-clade defined by the combined presence of a transversion to A at np 16114 and a transition at np 16274 that has also been observed in Senegalese Mandenka (Graven *et al.* 1995) and Wolof (Rando *et al.* 1998). L1b1 presents a TMRCA of about 36000 years (Figure 2), predating the diversity of L0a1 in Guineans. The matches in this cluster have a West African distribution well represented in Mandenka (haplotypes GB8 and GB20) and their frequency is highest in the Fulani-western and Senegal-eastern language groups (Table 2). GB23 and GB24 are widespread in Africa and are found in nearly all West African populations considered here (Salas *et al.* 2002). Another West African specific clade, L1c, is present at a relatively low frequency (0–8%) yet with high haplotype diversity in the Guiné-Bissau sample.

Haplogroups L2a–L2c are frequent in Senegambia (Table 2) and reveal signatures of a recent expansion from a limited number of founder haplotypes that are shared between populations of different linguistic affiliation. In contrast, haplotypes belonging to haplogroup L2d are represented by single individuals and do not show a common founder sequence (Figure 2). Fifteen out of 42 L2a haplotypes sampled in Guinea Bissau had matches elsewhere: West Africa (Cabo Verde, Brehm *et al.* 2002; Wolofs & Senegalese, Rando *et al.* 1998; Mandenka, Graven *et al.* 1995) but can also be found in East, South and North Africa. The geographic distribution of L2b and L2c haplotypes is largely restricted to West Africa. Not surprisingly most of the haplotype matches are with Cabo Verdeans, Wolof and Senegalese. L2c is the haplogroup that shows a higher extent of shared lineages: Cape Verde, Senegal Mandenka, mixed Senegalese and São Tomé. The last case is likely due to a recent gene flow from the Cape Verde Islands (Brehm *et al.* 2002). However, several L2 haplotypes observed in Guineans appeared as unspecific to other West African populations but shared matches with East and North Africans. This was the case for the Balanta (BLE) haplotype GB44

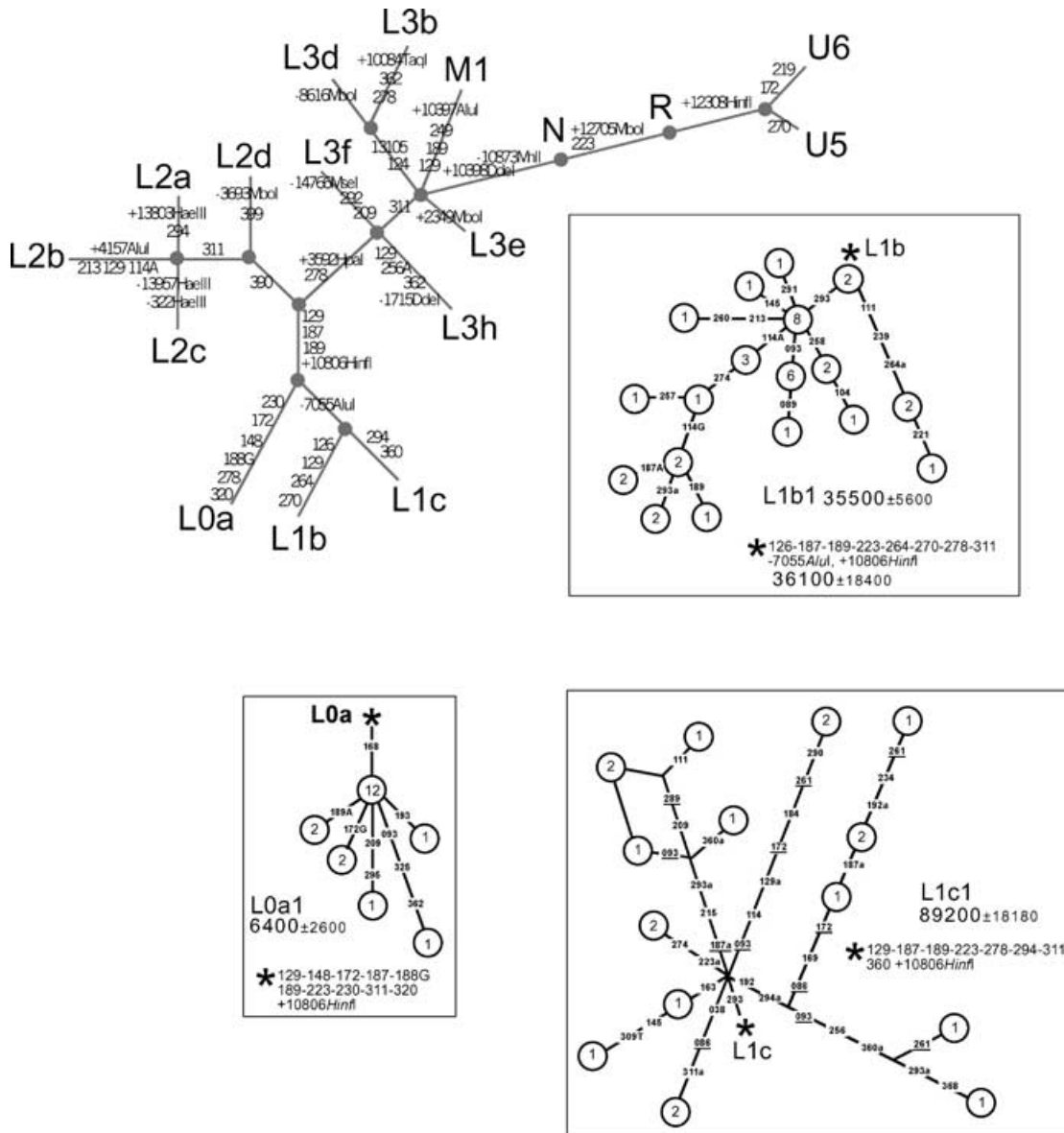


Figure 2 MtDNA phylogeny of all Guinean haplogroups and skeletons of various L0, L1, L2 and L3 sub-haplogroups based on HVS-I sequences and coding-region RFLPs. The number of individuals assigned to the haplotypes is shown within the circles. The numbers over the lines represent the HVS-I (-16000 bp) and coding region mutations, with respective restriction sites. Transversions are represented with suffixes (length variation in the C-stretch is disregarded). Recurrent mutations are underlined and *a* refers to the mutation loss relative to root. The star indicates the putative root of the haplogroup. Coalescence estimates \pm sd (in ya) are shown for haplogroups or sub-haplogroups.

matching only with Sudanese (Watson *et al.* 1999), and GB59 matching with Moroccan sequences. Interestingly, haplotype GB83 (L2b) found in the Mansonca (MSW) group had an exact match only with Ethiopians (our unpublished data). Also the Fula haplotype GB39 has not been reported in West Africa but appears in East Africa: Lake Turkana (Watson *et al.* 1997), Nubia,

Southern Sudan, Ethiopia and Saudi Arabia (our unpublished data).

Haplogroups L3b, L3d, and L3e are rare or absent in indigenous populations of North and South Africa but well represented in our sample. GB127 and GB134 are particular links of Guinean groups to North-west African Mozabites, Moroccans and Senegalese.

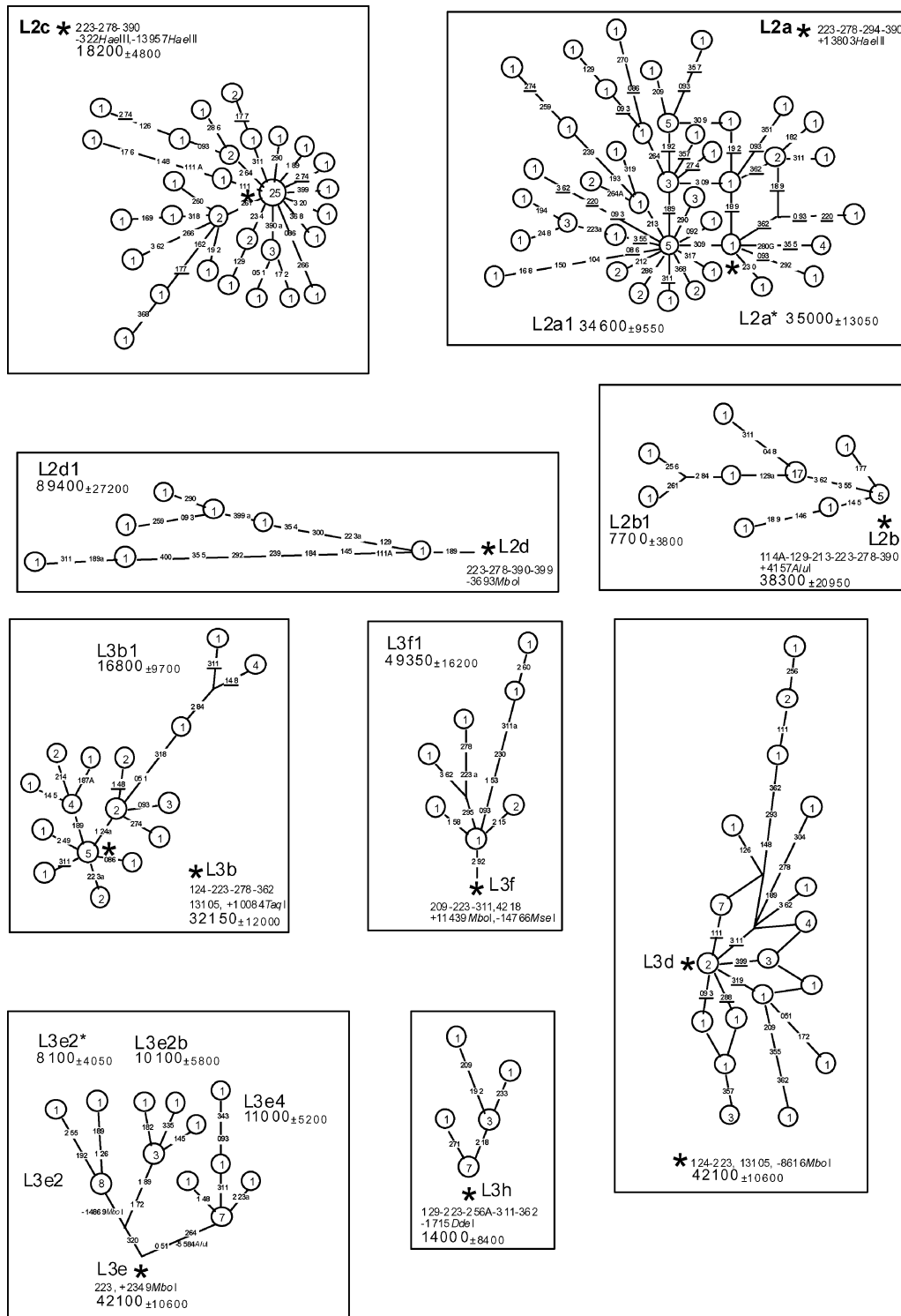


Figure 2 Continued.

Particularly, GB136 from Fula-related people has been found so far in Hausa and again in Nubians and Sudanese. Apart from Mozambique (6%) the majority of L3d lineages are West African (7% in mixed Senegalese

to 12% in Niger/Nigeria) with an estimated age of 42100 (± 10600, Salas *et al.* 2002). L3f is more frequent in Southeast Africa, ranging from 8% in Kenya/Sudan to 2% in Mozambique. The coalescence time of this

haplogroup in West Africa was calculated as 39400 ya (± 10400 , Salas *et al.* 2002), within the error range of the estimate based on Guinean samples (49350 ± 16200 ya). Haplotype GB178 in Fula shared an exact match with sequences from a wide range of East-African populations (Somalia, Egypt) and even Saudi Arabia. Haplogroup L3h is found in Ethiopia, Cape Verde and Niger/Nigeria at marginal frequencies ($\sim 1\%$) but reaches its highest known frequency in the Ejamat from Guinea (8%). Its coalescent time estimate (14000 ± 8400 ya) in Guineans is consistent with its late Pleistocene/early Holocene spread around Africa.

No significant differences between Guinean ethnic groups pooled by their linguistic affiliation were observed in haplogroup frequencies. As for their geographic neighbours (Table 2), haplogroups L1b, L1c, L2b, L2c, L2d, L3b, L3d, and L3e cover most of the mtDNA variation (64–85%). The Guiné-Bissau sample shows an overall genetic diversity of 0.901 (sd.005) that is significantly higher than among other samples from West Africa (Table 2).

M1 and U6 Lineages

Haplogroup M1 has been characterized as an East African remnant of the major Asian haplogroup M (Quintana-Murci *et al.* 1999). It has been found mostly in Ethiopian populations (17%), its characteristic HVS-I motif being also well represented in Egyptian and Sudanese populations along the Nile Valley (7–8%, Krings *et al.* 1999). HVS-I haplotypes matching the East African M1 clade have also been identified in Northwest Africans (Plaza *et al.* 2003, unpublished data) where their frequency can reach 12.8% in Algerians and 4% among Moroccan and Algerian Arabs and Berbers. M1 is generally absent from autochthonous West African populations but was found among Balanta, Baiote, and Djola groups speaking Niger Congo Atlantic Bak languages. The Guinean M1 haplotypes matched exactly one West Saharan (Rando *et al.* 1998), 2 Mozabites (Côte-Real *et al.* 1996), 2 Iranian and one Saudi Arabian sequence (unpublished data). This lineage derives from a particular cluster defined by a mutation at position 16185, which is also found in Ethiopia, Morocco and North African populations (Plaza *et al.* 2003, our unpublished results).

Haplogroup U6 is rather frequent in NW Africa, among Algerian Berbers, Moroccans and Mauritani-ans (Côte-Real *et al.* 1996; Rando *et al.* 1998; Plaza *et al.* 2003), but is rare or absent in western sub-Saharan Africans. Three different U6 haplotypes were observed in Fula, Mandenka and Manjaco groups. These haplotypes match with sequences of a wide geographic range: North and West Africa (Cabo Verde, Tuareg, Mozabites, Moroccan Arabs and Berbers), East Africa (Nile Valley, Egypt and Ethiopia), the Middle East (Iran) and Mediterranean Europe (Sicily and Portugal, <http://www.ahg.com/>), suggesting that their spread might be related to the southern expansions of the Berber groups to whom the Fulani languages relate.

European Lineages: U5

Ten individuals out of 372 samples, all related to Fulbe groups, carried mtDNA variants typical of western Eurasia, particularly Europe. Within these mtDNAs belonging to haplogroup U5 nine Fulanis share one particular HVS-I haplotype. Both haplotypes are only one mutational step away from a common node widespread in Europe. Although U5 is one of the most frequent mtDNA variants among western Eurasians (about 460 sequences in our mtDNA HVS-I database) no exact matches to the two Guinean haplotypes were found, as would be expected in the case of recent admixture. On the other hand, the Fulani U5 haplotype appears in a data set of West Africans (Wolof and Serer, Rando *et al.* 1998) and in Moroccans (unpublished data), pointing to the existence of a common African founder lineage of haplogroup U5. Again, as in haplogroup U6 the linguistic correlation suggests that the spread of the haplotype in Senegambia might be related to the movement of Berber populations. More data from North and West African populations is needed to better characterize the source and the time of the spread of this founder lineage.

AMOVA and Principal Component Analysis

Analysis of molecular variance (AMOVA) in African populations attributed 15.6% to differences between groups, 3% to variation between populations within groups, and 81.6% to differences within populations

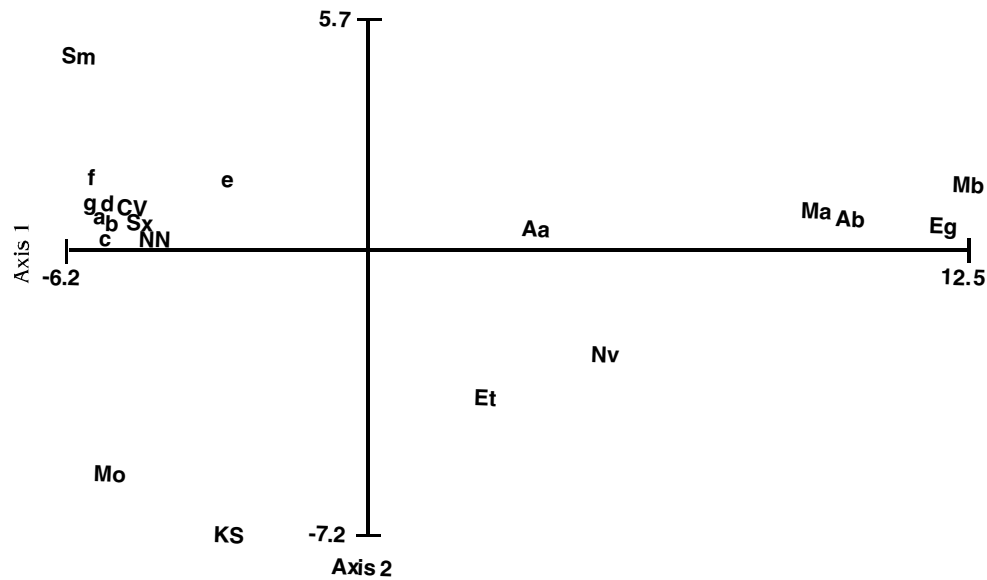


Figure 3 PCA of African populations based on data from Table 2 but excluding the !Kung/Khwe. Population codes are as follows: Mb (Morocco Berbers), Ab (Algerian Berbers), Ma (Morocco Arabs), Aa (Algerian Arabs), Eg (Egypt), Nv (Nile Valley), Et (Ethiopia), KS (Kenya/Sudan), Mo (Mozambique), Sm (Senegal Mandenka), Sx (Senegal mixed), NN (Niger/Nigeria), CV (Cabo Verde). Guinean ethnic groups were grouped (from a to g) as in Table 2. Axis 1 extracted 70.2% and axis 2, 12.5% of the total variation.

(overall $F_{ST} = 0.184$, $P < 0.0001$). A hierarchical structuring of populations into groups based on religion beliefs (Muslims vs. Animists) and geography (interior vs. littoral) gave similar values (data not shown).

A principal component (PC) analysis distinguished North Africans from sub-Saharan (Figure 3). The difference revealed by the first component is likely due to the presence of Eurasian mtDNA lineages among the North Africans and a relatively higher frequency of haplogroups L2a, L2c, L2d and U6 in Northwest Africa. The second component reflects L2/L0 frequencies. Moroccan Berbers and Arabs and Algerian Berbers are plotted close to Egyptians, supporting a common origin, while Algerian Arabs are placed apart. The Nile Valley sample occupies an intermediate position between Ethiopia and North Africans. The populations from Mozambique appear isolated and well differentiated from Kenya and Sudan. All the West Africans form a distinct and more compact cluster. Nevertheless the isolation of Senegalese Mandenka (Sm) and the Fula from Guiné (e) should be noted. As a whole, Guinean groups

are closer to West and then East Africa (see Axis 1, Figure 3).

Final Remarks

Roughly 87% of the mtDNA lineages found in the Guinean populations are common in other West African populations. Not surprisingly, the highest number of matches was with Cape Verde followed by other populations from the area (Mandenka, Wolof, Fulbe), but also with Morocco. The notable L haplotype sharing with North Africans testifies to the absence of a real barrier between this region and typical sub-Saharan populations. On the other hand, some Guinean groups (Fula and Balanta for instance) present haplotypes otherwise observed to date in East-African and Middle East populations.

It is interesting to note that the Bantu-associated markers L0a 9bp del CoII/tRNA^{Lys} (Soodyall *et al.* 1996), L3b motif 16124-16223-16278 (Watson *et al.* 1997), L3e1 particularly L3e1a characterized by mutation 16185 (Bandelt *et al.* 2001) or the 16192 L2a1

subclade (Pereira *et al.* 2001), were not found in our sample. This suggests that either Bantu migrations contributed very little to the gene pool of Guineans, despite the evidence of a Bantu migration starting from Cameroon and spreading towards Ghana, Nigeria, Burkina Faso and Mauritania, or that they had a distinct gene pool from that associated with the southwards migrants. The lack of Bantu branches of the Niger-Congo linguistic family, among a plethora of languages spoken in Guiné-Bissau, is more in agreement with the first hypothesis.

The finding of haplogroup M1 lineages of East African origin, albeit at low frequencies (3–5%) in Guinean groups with linguistic affinities to the Bak superfamily including Balanta, Baiote and Ejamat languages, supports the earlier suggestion of a Sudanese origin of the Balanta population and their spread to western Africa with kushitic migrants approximately 2000 years ago. Obviously, thereafter they were assimilated within the local population, acquiring their language. In particular the 16185 mutation might suggest a route through North Africa. The U6 presence in the Guinean pool, although at a low frequency, is not surprising, as these particular lineages have already been reported for this region. It seems plausible that the U5 lineages observed in the Fula arrived in Guiné via Sahel from North Africa before the slave trade. None of the typical European haplogroups (H, J, and T) were found in the present-day population of Guinea, whereas they exist at a fairly high frequency in North Africa in contrast to the U5 frequency (only 4.5%). This makes it less likely that the presence of U5 in Guiné, in particular, and in Northwest Africa in general, is due to recent admixture with the European population. A possible ancient migration from Asia to Africa was proposed by Cruciani *et al.* (2002) to explain the presence of some unusual Y-chromosome lineages identified in West Africa. Haplogroup R1 (defined by M173 mutation), without further branch defining mutations (M269 and M17) specific to Europeans, accounted for ~40% of the Y-chromosomes in North-Cameroon, while not yet having been sampled elsewhere in Africa. More data from Central and Western Africa are needed to cast light on the origin of such idiosyncratic mtDNA and Y chromosome lineages. Thus, our U5 sequences from the Guinean Fulbe people corroborate Cruciani's hypothesis of a prehistoric migration

from Eurasia to West Sub-Saharan Africa, testified by their present day restricted and localised distribution.

Acknowledgments

The authors are grateful for the precious help of the Chairman of the Joint Chiefs of Staff and the Ministry of Health from Guiné-Bissau. ICCTI (Lisbon, Portugal) and the Regional Government of Madeira provided financial support to AB. AR is a recipient of a Ph.D. scholarship from Fundação para a Ciência e Tecnologia (FCT, Lisbon) reference SFRH/BD/12173/2003. TK was supported by the Estonian basic research grant 4769. We are also grateful to the contributions from two anonymous reviewers to an early version of the manuscript.

Electronically Available Data

HVS-I and HVS-II haplotypes and their distribution among ethnic groups from Guiné-Bissau are available as Complementary Material at the web site <http://www.ahg.com/>. A list of the PCR primers and conditions used to amplify all pertinent mtDNA regions are also included in the Complementary Material web site.

References

- Alimen, H. (1987) Evolution du climat et des civilisations depuis 40000 ans du nord au sud du Sahara occidental. (Premières conceptions confrontées aux données récentes.) *Bull. L'Assoc. Franç. L'Étude Quaternaire* **4**, 215–227.
- Alves-Silva, J., Santos, M., Guimaraes, P., Ferreira, A. C., Bandelt, H.-J., Pena, S. D. & Prado, V. F. (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* **67**, 444–461.
- Almada, A. A. (1964) *Tratado breve dos rios da Guiné do Cabo Verde*, 2nd ed. Lisbon, 156p.
- Almeida, A. (1939) *Sobre a etno-economia da Guiné Portuguesa*. BGC, Lisboa Vol 15.
- Anderson, S., Bankier, A. T., Barrel, B. G., De Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
- Andrews, R. M., Kubacka, I., Hinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Hoewll, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics* **23**, 147.
- Aumassip, G., Ferhat, N., Heddouche, A. & Vernet, R. (1988) Le milieu saharien aux temps pré historiques. In: *Milieux*,

- hommes et techniques du Sahara pré historique. Problèmes actuels* (eds. Aumassip G. *et al.*), pp. 9–29. L'Harmattan (ed.), Paris.
- Bandelt, H.-J., Forster, P., Sykes, B. C. & Richards, M. B. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753.
- Bandelt, H.-J., Macaulay, V. & Richards, M. (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol Phylogen Evol* **16**, 8–28.
- Bandelt, H.-J., Alves-Silva, J., Guimarães, P., Santos, M., Brehm, A., Pereira, L., Coppa, A., Larruga, J. M., Rengo, C., Scozzari, R., Torroni, A., Prata, M. J., Amorim, A., Prado, V. F. & Pena, S. D. J. (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann Hum Genet* **65**, 549–563.
- Barros, A. (1947) A invasão fula da circunscrição de Bafatá. Queda dos Beafadas e Mandingas, tribos “Gabungabé”. *BCGP* (Bissau) **15**, 737–743.
- Brehm, A., Pereira, L., Bandelt, H.-J., Prata, M. J. & Amorim, A. (2002) Mitochondrial portrait of the Cabo Verde Archipelago: the Senegambian outpost of Atlantic slave trade. *Ann Hum Genet* **66**, 49–60.
- Camps-Fabrer, H. (1989) Capsien et Natoufien au Proche-Orient. In: *Colloque L'homme maghrebien et son environnement depuis 100.000 ans. Trav. Du CAPMO* (Algerie) 71–104.
- Carreira, A. (1962) O fundamento dos etnónimos na Guiné Portuguesa. *Revista Garcia da Horta* **10**, 305–323.
- Carreira, A. & Quintino F. R. (1964) Antroponímia da Guiné Portuguesa. *JIU*, Lisboa, Vol. 49 and 52.
- Carreira A. (1983) Migrações nas Ilhas de Cabo Verde. Instituto Caboverdeano do Livro. 2nd. Ed., Lisboa.
- Chen, Y.-S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K. & Wallace, D.C. (2000) mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am J Hum Genet* **66**, 1362–1383.
- Côrte-Real, H. B. S. M., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S. & Cambon-Tomsen, A. *et al.* (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331–350.
- Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., Wallace, D. C., Oefner, P. J., Torroni, A., Cavalli-Sforza, L. L., Scozzari, R. & Underhill, P. (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* **70**, 1197–1214.
- Diallo, T. (1972) Les institutions politiques du Fouta Djallon. *Initiation et Études Africaines*, IFAN, Dakar, vol 28.
- Dutour, O., Vernet, R. & Aumassip, G. (1988) Le peuplement pré historique du Sahara. In: *Milieux, hommes et techniques du Sahara pré historique. Problèmes actuels* (eds. Aumassip, G. *et al.*), pp. 39–52. L'Harmattan (ed.), Paris.
- Fage, J. (1995) *A history of Africa*. Routledge (ed.), London.
- Finnilä, S., Hassinen, I. E., Ala-Kokko, L. & Majamaa, K. (2000) Phylogenetic Network of the mtDNA Haplogroup U in Northern Finland Based on Sequence Analysis of the Complete Coding Region by Conformation-Sensitive Gel Electrophoresis. *Am J Hum Genet* **66**, 1017–1026.
- Forster, P., Harding, R., Torroni, A. & Bandelt, H.-J. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* **59**, 935–945.
- Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. & Excoffier, L. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* **12**, 334–345.
- Hair, P. E. (1967) Ethnolinguistic continuity on the Guinea Coast. *JAH*, Vol VIII, 247–268.
- Krings, M., Salem, A. E., Bauer, K., Geisert, H., Malek, A. K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., Sajantila, A., Pääbo, S. & Stoneking, M. (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? *Am J Hum Genet* **64**, 1166–1176.
- Lopes, C. (1999) Kaabunké, espaço, território e poder na Guiné-Bissau, Gâmbia e Casamance pré-coloniais. CND, Lisboa.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonnè-Tamir, B., Sykes, B. & Torroni, A. (1999) The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.
- Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., Sukernik, R. I., Olckers, A. & Wallace, D.C. (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* **100**, 171–176.
- Moreira, J. M. (1964) Os Fulas da Guiné Portuguesa na panorâmica geral do mundo Fula. *Boletim Cultural da Guiné Portuguesa* XIX, 417–432.
- Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.
- Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M. & Santachiara-Benerecetti, A. S. (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* **62**, 420–434.
- Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J. & Amorim A. (2001) Prehistoric and historic traces

- in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* **65**, 439–458.
- 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., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nature Genetics* **23**, 437–441.
- Quintino, F. (1964) Sobrevivências da Cultura etiópica no ocidente africano. *BCG (Guiné)* **19**, 5–35.
- Quintino, F. (1967) Os povos da Guiné. *BCG (Guiné)* **22**, 5–40.
- Quintino, F. (1969) Os povos da Guiné. *BCG (Guiné)* **24**, 861–915.
- Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H.-J. (1998) Mitochondrial DNA analysis of North-west African populations reveals genetic exchanges with European, Near-Eastern, and Sub-Saharan populations. *Ann Hum Genet* **62**, 531–550.
- Rando, J. C., Cabrera, V. M., Larruga, J. M., Hernández, M., González, A. M., Pinto, F. & Bandelt, H.-J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* **63**, 413–428.
- 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., Gölge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Nørby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Torroni, A. & Bandelt, H.-J. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.
- Richards, M. & Macaulay, V. (2001) The mitochondrial gene tree comes of age. *Am J Hum Genet* **68**, 1315–1320.
- Salas, A., Richards, M., De la Fé, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V. & Carracedo, A. (2002) The making of the African mtDNA landscape. *Am J Hum Genet* **71**, 1082–1111.
- Schneider, S., Roessli, D. & Excoffier, L. (2000) Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Soodyall, H., Vigilant, L., Hill, A. V., Stoneking, M. & Jenkins, T. (1996) mtDNA control-region sequence variation suggests multiple independent origins of an “Asian-specific” 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* **5**, 595–608.
- Stuhlmann, F. (1910) Handwerk und Industrie in Ostafrika, kulturgeschichtliche, Betrachtungen, Friedrichsen. Hamburg, 163p.
- Teixeira da Mota, A. (1954) Guiné Portuguesa. Agência Geral do Ultramar (ed.), Lisboa.
- Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus M. L. & Wallace D. C. (1997) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.
- Torroni, A., Rengo, R., Guida, V., Cruciani, F., Sellito, D., Coppa, A., Calderon, F. L., Simionati, B., Valle, G., Richards, M., Macaulay, V. & Scozzari, R. (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* **69**, 1348–1356.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A.C. (1991) African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507.
- Watson, E., Forster, P., Richards, M. & Bandelt, H.-J. (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* **61**, 691–704.
- Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisc. (1999).

Received: 2 July 2003

Accepted: 12 November 2003