Ann. Hum. Genet. (2001), **65**, 439–458
Printed in Great Britain

Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade

L. PEREIRA[1,2], V. MACAULAY[3], A. TORRONI[4,5], R. SCOZZARI[5], M.-J. PRATA[1,2] AND A. AMORIM[1,2]

[1] *Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias s/n, 4200 Porto, Portugal*
[2] *Faculdade de Ciências da Universidade do Porto, Pr. Gomes Teixeira, 4050 Porto, Portugal*
[3] *Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, UK*
[4] *Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy*
[5] *Dipartimento di Genetica e Biologia Molecolare, Università 'La Sapienza', Piazzale Aldo Moro 5, 00185 Roma, Italy*

(Received 3.4.01. Accepted 26.6.01)

SUMMARY

A sample of mitochondrial DNA (mtDNA) from the southeastern African population of Mozambique has been shown to have affinities with populations both to its north and south. From the north came sequences that may have been involved in the Bantu expansion (from western, through eastern, to southern Africa), such as members of haplogroups L3b, L3e1a and a subset of L1a. The dating of the major component of Mozambican mtDNAs, the subset L2a of haplogroup L2, displayed an age range compatible with the Bantu expansion. The southern influence was traced by the presence of sequence types from haplogroup L1d, a probable relict of Khoisan-speaking populations that inhabited the region prior to their displacement by the Bantu-speaking incomers. Within historical times, the forced displacement of Mozambicans as part of the slave trade, mainly documented as being to the Americas, generated a differential input of eastern African sequences into the mtDNA pools of the Americas and of Europe, as testified to by the greater number of sequence matches between Mozambique and the Americas, compared to those between Mozambique and Europe.

INTRODUCTION

In the last decade, data on African mtDNA have been accumulated with the aim of unravelling something of the demographic phenomena that have contributed to the settlement of the continent. This task is still at an early stage and is made potentially more difficult in Africa than in the rest of the world because the characteristic demographic phenomena of recurrent migration, population expansion and contraction including bottlenecks, population sub-structure generated by limits on gene flow, and more recent admixture effects, have occurred over a longer time depth. In addition, there is a rather poor archaeological context in which to place the genetics.

Several studies of restriction-fragment length polymorphisms (RFLPs) (Cann *et al.* 1987; Chen *et al.* 1995), of control region sequences (Vigilant *et al.* 1991; Soodyall, 1993; Krings *et al.* 1999) or a combination of both (Graven *et al.* 1995; Watson *et al.* 1997; Chen *et al.* 2000), have involved African populations. However, the sampling is still surprisingly patchy, and quite

Correspondence: Luísa Pereira, Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr Roberto Frias s/n, 4200 Porto, Portugal. Tel: +351 22 5570700; Fax: +351 22 5570799. E-mail: lpereira@ipatimup.pt

poor in the southeast. Besides the intrinsic interest of seeing how this region fits into the emerging picture of African mtDNA diversity, it has potential implications on a wider scale too. South-east Africa was an important source of slaves from 1643 onwards, when individuals from Mozambique and Madagascar constituted a major proportion of the slaves shipped by the Portuguese to the former European colonies in America, e.g. Brazil and the Caribbean, in such a way that 'by the eighteenth century this commerce, directed to the Americas, was more important on that coast than anywhere else' (Thomas, 1998). Records point to ~ 1000000 slaves originating from Mozambique/Madagascar in a total of ~ 13000000 leaving African ports (Thomas, 1998). By that time, slave importation was already reduced in Europe, and the majority of European countries forbade entry of black slaves by the middle of the eighteenth century. So the eastern sub-Saharan contribution to the European African sequences, sporadically detected in some countries of this continent (Côrte-Real *et al.* 1996; Pereira *et al.* 2000), is expected to have been reduced compared to its influence in America.

Here, we present hypervariable region I (HVRI) and II (HVRII) data for Mozambique, a south-east African population which was a Portuguese colony between 1752 and 1975. This country contains several ethnic groups, nearly all Bantu-speaking. Originally a linguistic classification, referring to a widespread group of languages within the South-Central Niger-Congo family, Bantu now refers to a complex of physical, anthropological and genetic characters correlated with the linguistic distribution, which are explained by a large-scale Holocene range expansion. This expansion occurred in several waves and directions and was responsible for the dispersal of farming to southern and central Africa. The linguistic evidence points to a Bantu origin in the vicinity of the Cross River valley near the present-day border between Nigeria and Cameroon (Newman, 1995). Around 5000 years ago, the Bantu expansion began in two directions: south-western, arriving at the equatorial rain forest 3500 years ago, and eastern, entering the fringes of the interlacustrine region in what is now Uganda, 3000 years ago, forming the eastern Bantu core area. From this new core, two new expansions moved towards South Africa: one group along the Ruvuma River toward the coast, reaching present-day Natal by the end of the third century A.D., and the other along the shores of Lake Malawi, through what is now eastern Zimbabwe, reaching the northern Transvaal around A.D. 500. At the mitochondrial level some Bantu expansion markers have been proposed: Soodyall *et al.* (1996) and Watson *et al.* (1997) pointed to a 9-bp deleted subset of haplogroup L1a and to the haplogroup L3b, while Bandelt *et al.* (in press) proposed that haplogroup L3e1a must have been prominent in the southern Bantu expansion. However, particularly in the case of L1a, a detailed dissection of the phylogeography, with a well-resolved phylogeny, has still to be performed: the signal could easily be of an earlier Holocene event.

Our aims here are to describe the phylogeography of Mozambique sequences in the context of African variation, and to search for possible sequence matches outside Africa that might shed light on the slave trade.

Material and methods

Subjects

A total of 109 unrelated individuals born in Mozambique were analysed and DNA was extracted from blood spots by the resin Chelex-100 method (Lareu *et al.* 1994). Individuals belonged to different ethnic groups (Changana, 35; Ronga, 21; Chope, 12; Bitonga, 8; and Matsua, 8; and 25 to various other groups), but all were Bantu speakers (http://www.sil.org/ethnologue/countries/Moza.html).

HVRI and II amplification and sequencing

Mitochondrial DNA was amplified using the primers L15997 (5′-CACCATTAGCACCCAAAGCT-3′) and H16401 (5′-TGATTTCACGGAGGA-

Table 1. *Code, place of origin, sample size and bibliographic references for the populations studied*

Code	Place (ethnic group)	Sample size	Reference
Northern African			
SAH	Western Sahara	25	Rando *et al.* (1998)
MA	Mauritania	30	Rando *et al.* (1998)
MO	Morocco	32	Rando *et al.* (1998)
BM	Morocco (Berber)	60	Rando *et al.* (1998)
EGY	Egypt	68	Krings *et al.* (1999)
MZB	Algeria (Mozabite)	86	Côrte-Real *et al.* (1996); Macaulay et al. (1999)
Western African			
HA+KA	Niger (Hausa and Kanuri)	20+14	Watson *et al.* (1997)
FUL	Nigeria (Fulbe)	60	Watson *et al.* (1997)
SON+TU	Nigeria (Songhai and Tuareg)	10+23	Watson *et al.* (1997)
YOR	Nigeria (Yoruba)	21+14	Watson *et al.* (1997); Vigilant *et al.* (1991)
SEN	Senegal	50	Rando *et al.* (1998)
SER	Senegal (Serer)	23	Rando *et al.* (1998)
WO	Senegal (Wolof)	48	Rando *et al.* (1998)
MAN	Senegal (Mandenka)	119	Graven *et al.* (1995)
Central African			
MBU	Zaire (Mbuti)	20	Vigilant *et al.* (1991)
BIA	Central African Republic (Biaka)	17	Vigilant *et al.* (1991)
Eastern African			
TK	Kenya (Turkana)	36	Watson *et al.* (1997)
SO	Somalia	27	Watson *et al.* (1997)
KIK	Kenya (Kikuyu)	25	Watson *et al.* (1997)
NUB	Nubia	80	Krings *et al.* (1999)
SUD	Southern Sudan	76	Krings *et al.* (1999)
South-eastern African			
MOZ	Mozambique	109	This work
Southern African			
KNG1	Botswana (!Kung)	25	Vigilant *et al.* (1991)
KNG2	South Africa (!Kung)	43	Chen *et al.* (2000)
KWE	South Africa (Khwe)	31	Chen *et al.* (2000)
HER	South Africa (Herero)	26	Vigilant *et al.* (1991)

TGGTG-3′) for HVRI and L48 (5′-CTCACGG-GAGCTCTCCATGC-3′) and H408 (5′-CTGTTA-AAAGTGCATACCGCCA-3′) for HVRII. The temperature profile was 95 °C for 10 s, 60 °C for 30 s and 72 °C for 30 s, for 35 cycles of amplification. The amplified samples were purified with Microspin™ S-300 HR columns (Pharmacia Biotech), according to the manufacturer's specifications. The sequence reactions were carried out using the kit Big-Dye™ Terminator Cycle Sequencing Ready Reaction (Perkin–Elmer), with one of the above primers, in both forward and reverse directions. A protocol based on $MgCl_2$/ethanol precipitation was used for post-sequence reaction purification of samples, which were then applied to a 6% PAGE and run in an automatic sequencer ABI 377.

RFLP analyses of haplogroup L3 sequences

In order to check the assignment of a number of sequence types to haplogroup L3 and its subclusters, we checked the following RFLPs in putative members of L3: 2349*Mbo*I (present in L3e), 3592*Hpa*I (absent in L3 in general), 8616*Mbo*I (absent in L3d) and 10084*Taq*I (present in L3b), in the numbering system of the Cambridge Reference Sequence (CRS) of Anderson *et al.* (1981). PCR amplifications were performed using primers and conditions described by Torroni *et al.* (1992). Digestions were carried out according to the manufacturer's specifications and the resulting fragments were run in 9% polyacrylamide gels and visualized by silver staining (Budowle *et al.* 1991).

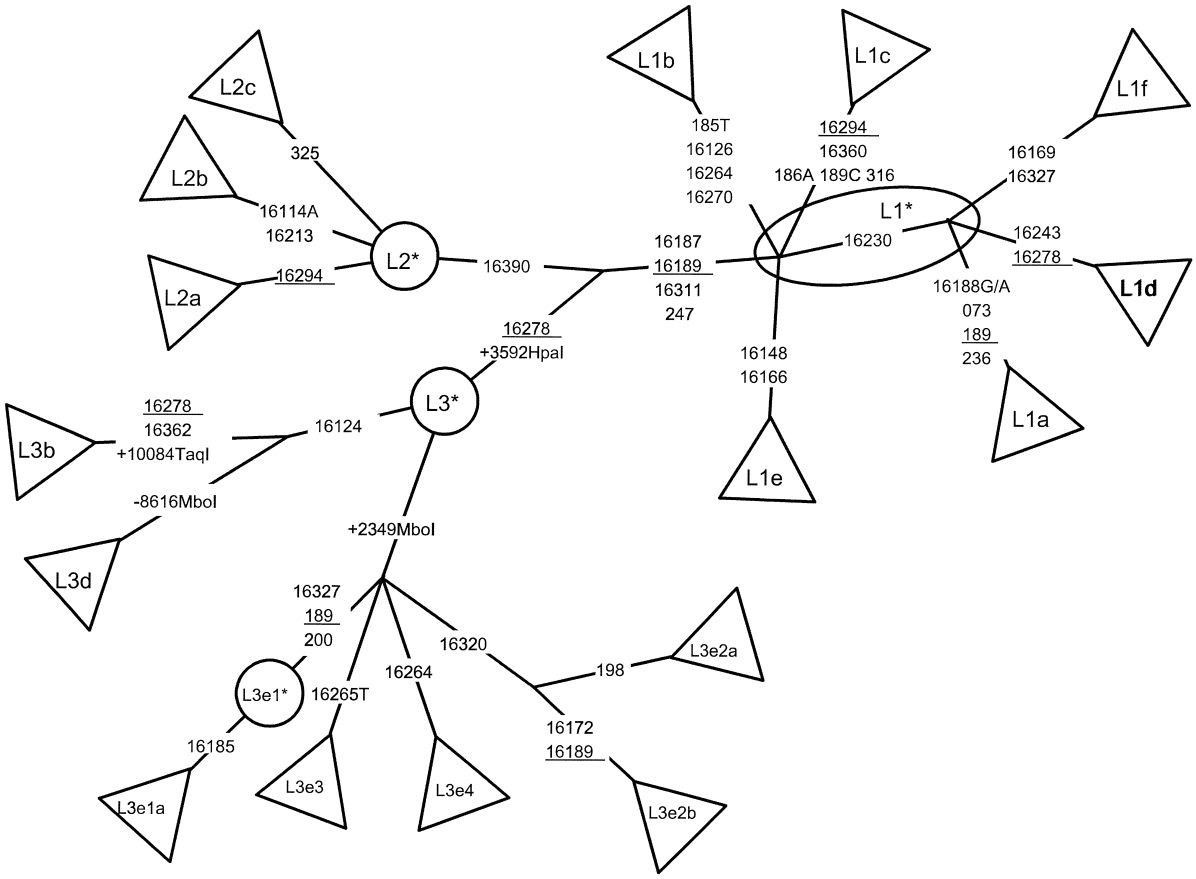

Fig. 1. A schematic phylogeny of African haplogroups used in our classification of sequences from Mozambique. We have drawn on information from Watson *et al.* (1997), Chen *et al.* (2000) (although our naming scheme is different from theirs), Alves-Silva *et al.* (2000), Richards & Macaulay (2000), Ingman *et al.* (2000), Bandelt *et al.* (in press) and also on some unpublished information. Triangles represent well-characterized clades in the mtDNA phylogeny. Circles and ellipses correspond to possibly paraphyletic groupings of less well-characterized haplotypes. The node marked L3* has the motif 16223T, 073G, 263G in the control region with respect to the CRS. Mutations (shown on the branches) are transitions unless a nucleotide is specified. Underlining of a position indicates that it mutates more than once in the figure. Four diagnostic RFLPs are also shown since they were checked in a number of samples; the direction of site loss and gain is indicated with respect to the node L3*. We keep the existing L1 nomenclature despite the fact that L1 is not a clade: it harbours the root of the tree, which occurs on a branch separating L1a and certain lineages present in Khoisan-speaking populations from the rest of the tree (Ingman *et al.* 2000). There are still uncertainties in the phylogeny, e.g., whether the 16124 mutation occurs independently in L3b and L3d: however these ambiguities did not affect the classification of the haplotypes reported here.

Genetic analysis and population comparison

The nucleotide positions considered for the analysis of the sample were 16024–16383 for HVRI and 73–340 for HVRII. Length variation (often scored as transversions in HVRI) was not considered in the analysis (Bendall & Sykes, 1995).

Sequence classification into haplogroups was according to Watson *et al.* (1997), Rando *et al.* (1998), Macaulay *et al.* (1999), Alves-Silva *et al.* (2000), Richards & Macaulay (2000) and Bandelt *et al.* (in press). Figure 1 displays a phylogeny of the African haplogroups used to classify the haplotypes. Given the relatively high mutation rate of mtDNA, especially in the control region, and the large time depth of some of the African haplogroups, recurrent mutations at motif positions occur quite often. It is necessary to be sensitive to this possibility when classifying haplotypes.

The Mozambique data were compared with data from several populations from various parts of Africa. The code, place of origin, sample size and bibliographic references of these population samples are displayed in Table 1. In some samples for which there were few individuals, we combined neighbouring populations or the same population from different studies. For this analysis, we only considered HVRI from nucleotide positions 16090–16365.

Molecular diversity indexes, mismatch distributions, analysis of variance (AMOVA) and tests of the standard neutral model (via Tajima's D and Fu's F_s statistics) were calculated in ARLEQUIN 2.0 (Schneider *et al.* 2000). Principal component (PC) analyses were performed using POPSTR (H. Harpending, pers. comm.) on the haplogroup composition of the various African populations. For these analyses we used the frequencies of the following haplogroups: pre-HV, N1a together with N1b (Richards *et al.* 2000), I, J, K, T, U6, the rest of U, X (Macaulay *et al.* 1999), M1 (Quintana-Murci *et al.* 1999), L1a, L1b, L1c, L1d, L1e, L1f, L1* (consisting of non-L2/L3 types not classified as L1a-f), L2a, L2b, L2c together with L2* (since these cannot be distinguished without HVRII information), L3*, L3b, L3d, L3e1, L3e2, L3e3, L3e4 (Figure 1) and 'other'. A population neighbour-joining tree was obtained using PHYLIP (Felsenstein, 1993) from pairwise F_{ST} values estimated in ARLEQUIN 2.0. These F_{ST} values incorporate information both on haplotype frequencies and the genetic distances between haplotypes, calculated as the number of nucleotide positions differing between pairs of sequences. Reduced median networks were constructed by hand and checked in NETWORK 2.0d (Bandelt *et al.* 1995). The dates of the most recent common ancestor of specific subclusters in the phylogeny were estimated using ρ, the average number of transitions from the ancestral sequence type to all sequences in the cluster, in conjunction with a mutation rate estimate of 20180 years per transition in the sequence stretch 16090–16365 (Forster *et al.* 1996). Standard errors were calculated as in Saillard *et al.* (2000).

RESULTS AND DISCUSSION

Sequencing and RFLP results

The HVRI and HVRII sequences obtained and the results of the selective RFLP typing are shown in the Appendix.

HVRI and HVRII diversity in Mozambique

Diversity measures for both hypervariable regions are displayed in Table 2. As has been described previously, HVRII is rather less diverse than HVRI (Pereira *et al.* 2000). In both segments, the diversity in Mozambique is higher than in a typical European population, i.e. the mean pairwise difference is 1.5–2 times greater.

Mismatch distributions (Fig. 2) in Mozambique were very ragged for both HVRI and HVRII, which is consistent with results for other sub-Saharan African populations (Bandelt & Forster, 1997). An interesting point was the considerable number of identical sequences (especially for HVRII), which could point to a sample bias effect of some ethnic groups consisting of closely related individuals. However this does not seem to be the case, since most of the identical sequences belong to different ethnic groups.

We further investigated if there was any substructuring of mtDNA between individuals from the Mozambique ethnic groups for which we had a substantial number of samples. The application of AMOVA showed that there was no evidence of significant variation between ethnic groups ($p = 0.60$).

Comparison with other African populations

In order to set the diversity observed in Mozambique within a continental context, we compared this population with a database of dispersed African populations. Since data for HVRII are scarcer we considered only HVRI diversity.

We used several methods, some based on haplogroup frequencies, such as PC analysis, and

Table 2. *Diversity measures in Mozambique within HVRI and HVRII*

	Haplotypes[1]	Segregating sites[2]	Gene diversity[3]	Mean pairwise difference
HVRI	50 (45.9)	60 (16.7)	0.962 ± 0.008	7.86
HVRII	35 (32.1)	29 (10.8)	0.846 ± 0.032	5.18
HVRI + HVRII	64 (58.7)	89 (14.1)	0.973 ± 0.007	13.04

[1] Number of distinct haplotypes in sample (percentage of sample size).
[2] Number of sites variable in sample (percentage of all sites).
[3] Average heterozygosity ± standard error.

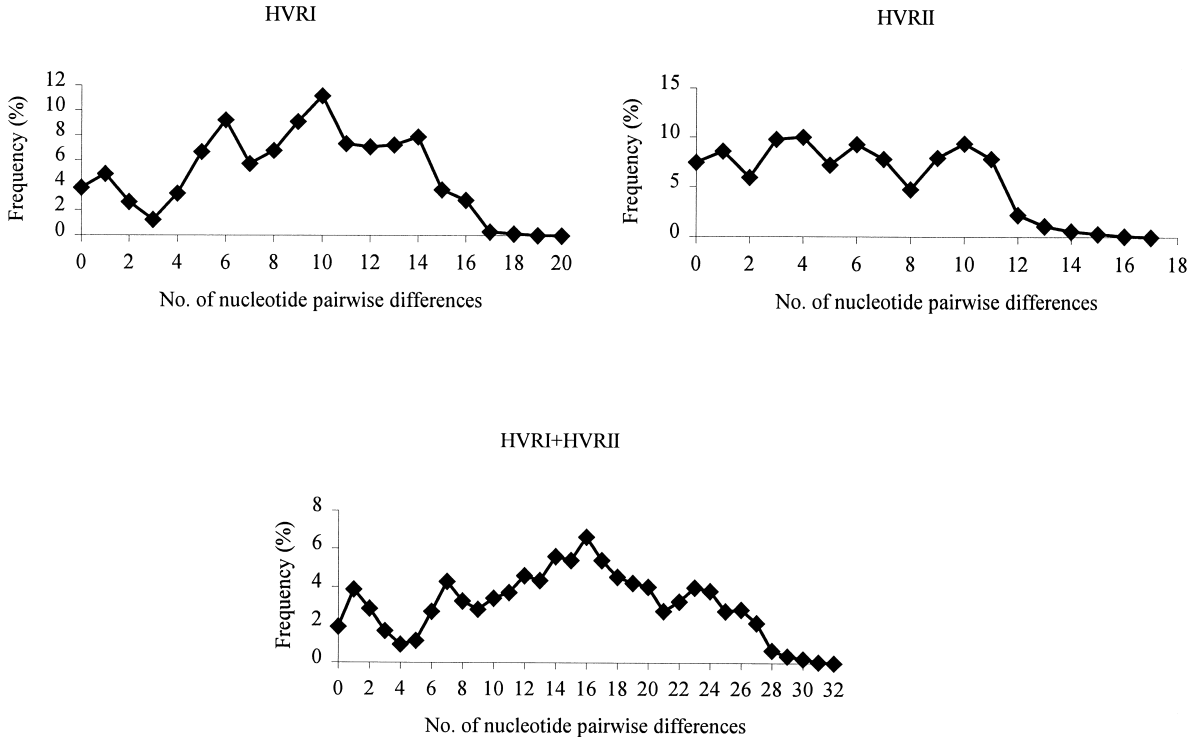

Fig. 2. Mismatch distributions for HVRI and HVRII in Mozambique.

others based on sequence diversity, such as AMOVA.

Molecular diversity

Overall diversity for HVRI in Africa (Table 3) is highest for western and eastern populations, followed closely by northern ones. The two Pygmy groups (here collectively referred to as Central African) and southern Africans displayed considerably lower diversity.

Departure from the standard (null) model of populations evolving at constant size in mutation-drift equilibrium with no selection was tested by employing Tajima's D and Fu's F_s statistics. For Tajima's D there were significant negative values only in northern populations and in Sudan. All the other populations showed non-significant values, and some southern (Khoisan-speaking) populations and the Pygmies even displayed positive values (although also non-significant). For Fu's more powerful F_s statistic, there were more populations with significant negative values, not only restricted to the north, but also present in the west and the east, while non-significant positive values occurred as for D. If we neglect the hypothesis of selection, the significant negative D and F_s values provide an indication of population expansion in most populations except in the Pygmies and Khoisan. This observation does not necessarily imply absence of expansion in the Pygmies and Khoisan, but possibly that the signal was lost by

Table 3. *HVRI* (*from np 16090 to 16365*) *diversity and neutrality measures in African populations*

Population	Sample size	Haplotypes[1]	Segregating sites[2]	Gene diversity[3]	Mean pairwise differences	D[4]	F_s[5]
Northern African							
SAH	25	20 (80.0)	29 (10.5)	0.973±0.022	5.11	−1.25	−12.4***
MA	30	22 (73.3)	28 (10.1)	0.970±0.018	5.83	−0.63	−11.5***
MO	32	29 (90.6)	44 (15.9)	0.988±0.014	5.84	−1.70*	−25.0***
BM	60	38 (63.3)	47 (17.0)	0.963±0.015	4.44	−1.88**	−25.7***
EGY	68	59 (86.8)	66 (23.9)	0.993±0.005	6.82	−1.70**	−25.1***
MZB	85[6]	29 (34.1)	35 (12.7)	0.942±0.010	4.73	−1.02	−11.1***
Western African							
HA+KA	34	31 (91.2)	41 (14.9)	0.995±0.009	6.19	−1.38	−25.2***
FUL	60	38 (63.3)	43 (15.6)	0.972±0.010	6.82	−0.87	−23.2***
SON+TU	33	29 (87.9)	41 (14.9)	0.992±0.009	7.26	−1.02	−21.3***
YOR	34[6]	32 (94.1)	44 (15.9)	0.996±0.008	7.31	−1.16	−25.0***
SEN	50	42 (84.0)	41 (14.9)	0.989±0.008	6.24	−1.08	−25.2***
SER	23	21 (91.3)	40 (14.5)	0.992±0.015	8.09	−0.98	−12.0***
WO	48	39 (81.3)	42 (15.2)	0.991±0.006	7.50	−0.71	−25.0***
MAN	110[6]	46 (41.8)	47 (17.0)	0.963±0.008	6.23	−0.94	−24.5***
Central African							
MBU	13[6]	5 (38.5)	19 (6.9)	0.756±0.097	7.13	0.70	3.8
BIA	17	8 (47.1)	20 (7.2)	0.890±0.043	7.81	1.27	1.7
Eastern African							
TK	36	32 (88.9)	54 (19.6)	0.991±0.010	9.66	−0.94	−20.8***
SO	27	24 (88.9)	41 (14.9)	0.992±0.013	6.90	−1.32	−16.3***
KIK	25	23 (92.0)	45 (16.3)	0.993±0.013	7.96	−1.27	−14.6***
NUB	80	50 (62.5)	64 (23.2)	0.974±0.008	7.88	−1.29	−24.8***
SUD	76	63 (82.9)	73 (26.4)	0.993±0.004	8.33	−1.47*	−24.8***
MOZ	109	49 (45.0)	57 (20.7)	0.960±0.008	7.78	−0.89	−23.6***
Southern African							
KNG1	24[6]	9 (37.5)	16 (5.8)	0.830±0.053	2.97	−1.10	−1.3
KNG2	43	12 (27.9)	31 (11.2)	0.812±0.045	7.30	0.07	1.8
KWE	31	10 (32.3)	34 (12.3)	0.884±0.029	8.75	0.10	3.0

[1] Number of distinct haplotypes in sample (percentage of sample size).
[2] Number of sites variable in sample (percentage of all sites).
[3] Average heterozygosity±standard error.
[4] Tajima's D statistic (p-value: * = $0.01 < p \leqslant 0.05$; ** = $0.001 < p \leqslant 0.01$; *** = $p \leqslant 0.001$).
[5] Fu's F_s statistic (p-value: * = $0.01 < p \leqslant 0.05$; ** = $0.001 < p \leqslant 0.01$; *** = $p \leqslant 0.001$).
[6] Some sequences were not considered for this analysis since there were many positions not scored.

subsequent contractions (Bandelt & Forster, 1997).

AMOVA analysis

We grouped the different populations into large geographic zones: northern, western, eastern, central and southern, as displayed in Table 1, and investigated how the proportion of variance was distributed between groups and between populations in the same group, by AMOVA. When Mozambique was assigned to the eastern group, the proportion of variance between populations in the same group took its minimum value (4.9%, compared to 5.5% when included in the central group, 6.5% when included in the southern group and 5.2% when included in the western group, all values significantly greater than zero at the 5% level). In this case, the proportion of variance between groups was 11.8%. This suggests that our Mozambique sample may have closest affinities with the populations to its north.

NJ tree

A population NJ tree constructed from pairwise F_{ST}s (not shown) revealed the geographic clustering of the different populations, with a Khoisan/Pygmy cluster, a northern African cluster and a western African cluster. The eastern populations fall between the Khoisan/Pygmy

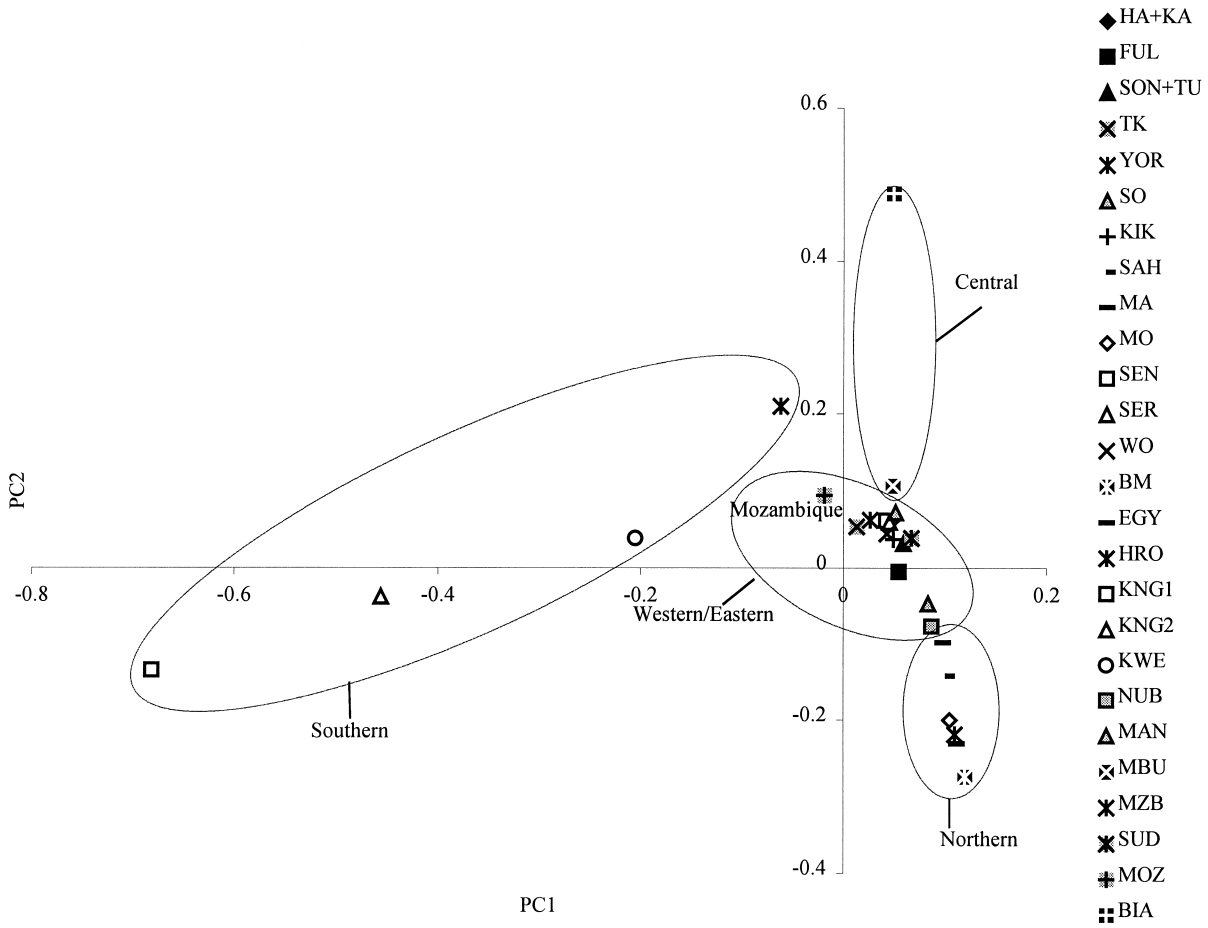

Fig. 3. The first two principal components of haplogroup frequency profiles for all African populations.

cluster and the rest. Mozambique clusters with western populations, although the branch that links it to them is long, in contrast to the short branches connecting western populations. As in the AMOVA analysis, the Mozambique population appears less like those populations to the south than to the north. A tree-like model of population evolution is unlikely to capture much of the reality of population history, so we proceeded with an exploratory PC analysis.

Principal component analysis

A preliminary PC analysis of all the populations plus a sample of Herero (Vigilant *et al.* 1991) (for which there was enough information for haplogroup classification, but too many uncharacterized sites for its inclusion in the previous analyses) showed some consistent geographical clustering, although the western and eastern populations were intermingled (Fig. 3).

The first two principal components amount to only 38% of the variation, leaving the rest uncharacterised. However, the first principal component (PC1), responsible for 22% of the variance, splits northern, western and eastern populations from southern ones and Mozambique. The main haplogroup responsible for this PC is L1d, which is typical of Khoisan populations (Bandelt & Forster, 1997); its presence in non-Khoisan populations may represent recent admixture. L1d constitutes 7% of the Mozambique sample and was absent in all the other non-southern populations analysed here (except one individual in the Turkana). The second principal component (PC2), responsible for 16% of the variance, distinguishes the northern African populations. The main haplogroups responsible for this PC are L1c and pre-HV. The pattern in L1c probably reflects its extremely high frequency in the Biaka (probably

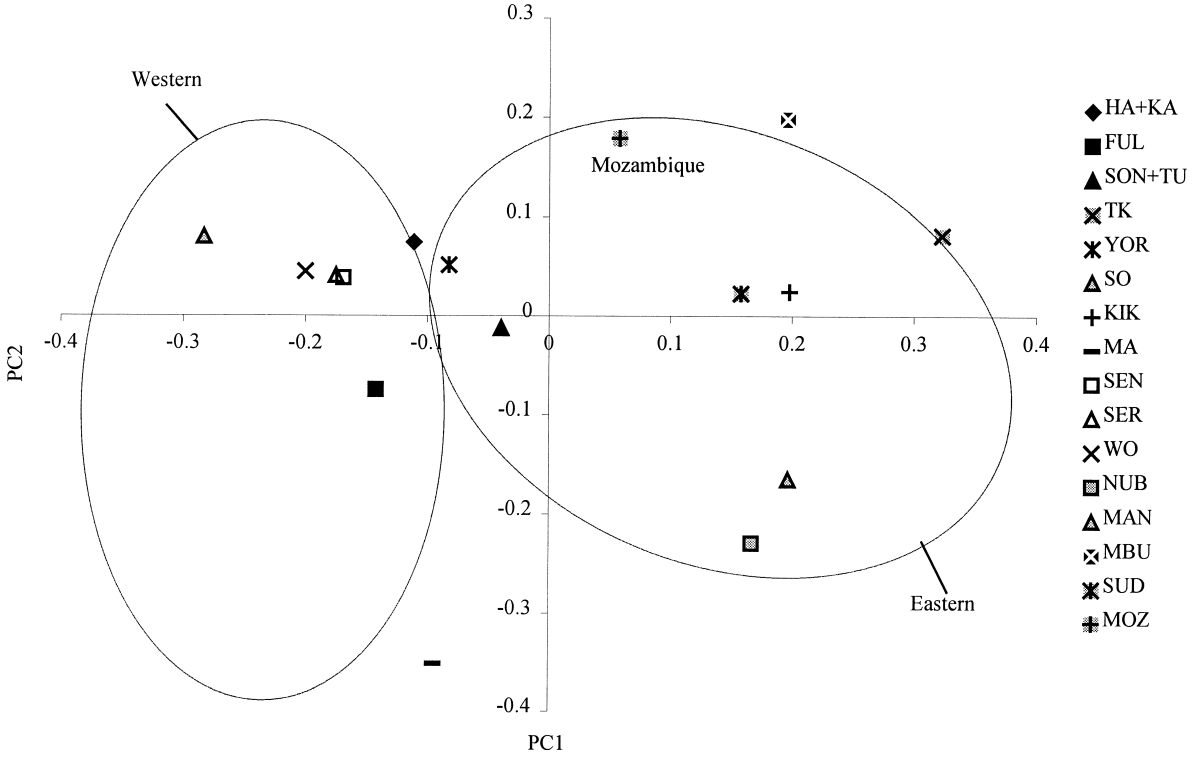

Fig. 4. The first two principal components of haplogroup frequency profiles for western and eastern African populations, including the Mbuti.

Table 4. *Number of sampled individuals in Mozambique by haplogroup* (*and frequency with standard error*)

Haplogroup	Frequency (%)
L1a	16 (14.7±3.4)
L1b	1 (0.9±0.9)
L1c	5 (4.6±2.0)
L1d	8 (7.3±2.5)
L1e	2 (1.8±1.3)
L2a	47 (43.1±4.7)
L2b	2 (1.8±1.3)
L2c	1 (0.9±0.9)
L3*	2 (1.8±1.3)
L3b	4 (3.7±1.8)
L3d	2 (1.8±1.3)
L3e1*	9 (8.3±2.6)
L3e1a	4 (3.7±1.8)
L3e2a	1 (0.9±0.9)
L3e2b	2 (1.8±1.3)
L3e3	2 (1.8±1.3)
L3e4	1 (0.9±0.9)

due to drift) compared to its near absence in the north, while that in the predominantly western Eurasian haplogroup pre-HV is accounted for by its virtual absence south of the Sahara.

In order to remove these large signals, which are not especially informative with regard to Mozambique, we performed a refined PC analysis by excluding the peripheral populations in Figure 3. Whereas in the previous analysis western and eastern populations were mixed up, the new PC1 (Fig. 4), responsible for 30% of the variance, splits the eastern populations and Mozambique from the western populations. Several haplogroups have a similar contribution: L3*, L1a, L1b, L2*+L2c and L1e. L3*, L1a and L1e are typically eastern haplogroups, and L1b and L2*+L2c are western haplogroups. The new PC2, responsible for 18% of the variance, splits the eastern and western populations in a north-south axis, and the main haplogroup responsible for this is pre-HV, as above.

Analyses and dating of sequence types

All the Mozambican sequences belong to sub-Saharan haplogroups (see Appendix). Past European (especially Portuguese) contact was not detected at the mtDNA level: there is a complete absence of European sequences (Richards *et al*. 2000). In addition, no east (Horai

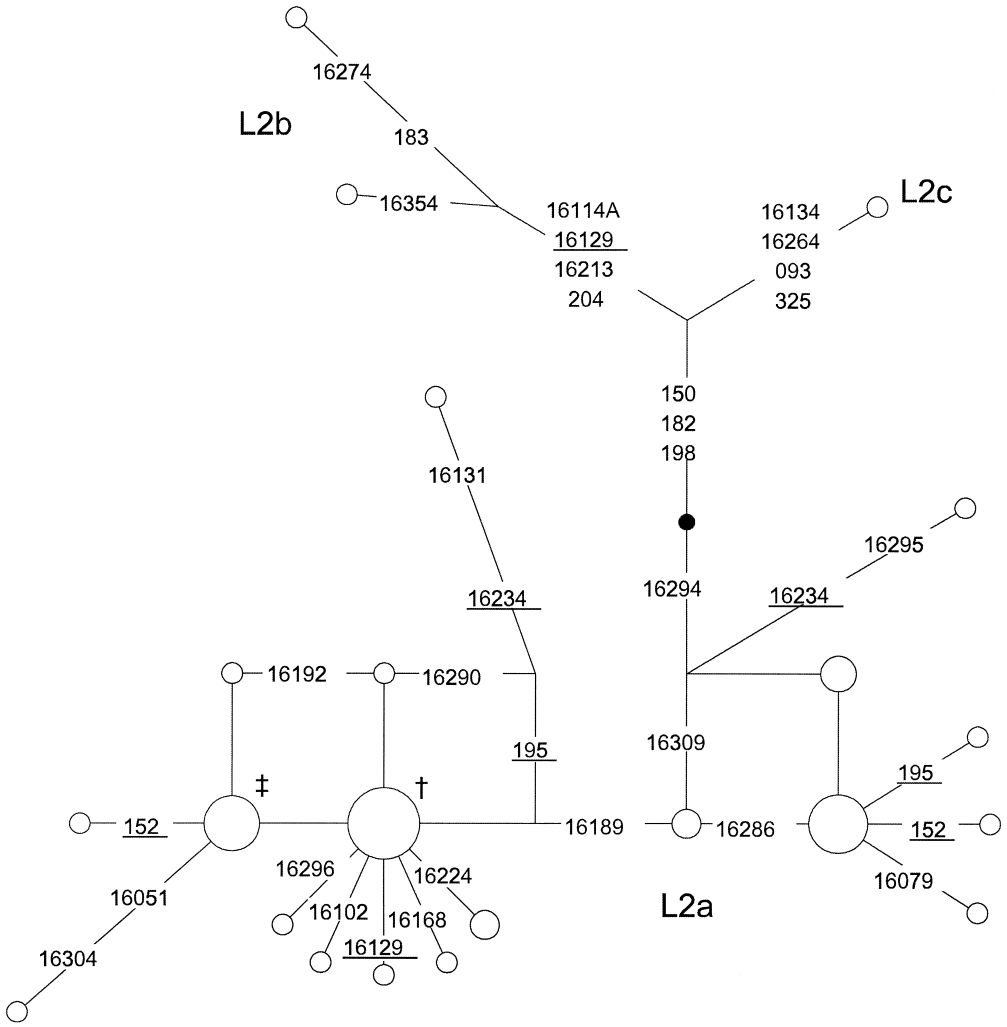

Fig. 5. The reduced median network of the 50 L2 sequences in the Mozambique sample. The circles are combined HVRI/HVRII haplotypes, the areas of which are proportional to the frequency in the sample. The smallest circles are singletons, the largest has frequency 12. Mutations (shown on the branches) are transitions unless a base change is explicitly indicated. Underlining indicates resolved recurrent mutations, unresolved events being shown by reticulation. The solid node (not observed in the sample) has the motif 16223–16278–16390–073–146–152–195–263. Branches are shown compressed in L2b and L2c for convenience. Two putative founder sequences are indicated by the symbols † and ‡.

et al. 1996) or south (Kivisild *et al*. 1999) Asian mtDNAs were detected. No Near Eastern sequences (Richards *et al*. 2000), detectable in some northeastern African populations, were observed in Mozambique, and the north African haplogroup U6 (Côrte-Real *et al*. 1996; Macaulay *et al*. 1999) was also absent.

Certain haplogroups were present at high frequency in the Mozambique sample (Table 4) and for these we performed a phylogenetic network analysis and examined their distributions across Africa, in an attempt to determine when they arrived in Mozambique.

Haplogroups L1a and L1d

Haplogroup L1a appears in our sample in two clusters of haplotypes. These clusters are inferred to correlate with the presence/absence of one instance of the intergenic COII/tRNA^{Lys} 9bp deletion (Soodyall *et al*. 1996). The non-deleted L1a types (usually with 16168T and 185A: Ingman *et al*. 2000) represent 10 mtDNAs in our sample (9%), while the six remaining L1a mtDNAs (6%) are likely 9bp-deleted.

Haplogroup L1d is by far the most common in Khoisan-speaking populations and, apart from

Table 5. *Haplogroup L2a diversity in Africa*

	Heterozygosity	Frequency (%)	Mean pairwise difference
Eastern African	0.953±0.023	20.5	3.82
Western African	0.952±0.013	19.0	3.14
Northern African	0.910±0.068	4.3	2.36
Mozambique	0.832±0.031	43.1	2.24
Central African	0.733±0.155	29.7	2.27
Southern African	0.000±0.000	3.0	0.00

one individual from the Turkana with an outlying L1d type (Watson *et al*. 1997), has so far not been observed north of Namibia. It is an early branch in the phylogeny, although its precise location is still subject to some uncertainty (compare Chen *et al*. 2000 with Watson *et al*. 1997). It is present in eight individuals (7%) of the Mozambique sample, who display seven different HVRI/HVRII haplotypes. These types could well represent a relict of the populations that inhabited this area before the Bantu and earlier migrations, although recent gene flow with Khoisan-speaking populations (Bandelt & Forster, 1997) cannot be excluded.

Haplogroup L2a

Figure 5 shows a reduced median network of haplogroup L2 in Mozambique. The majority (95%) of Mozambican L2 belong to the subcluster L2a, defined in HVRI by a transition at 16294 in addition to the L2 motif. This subcluster is widely distributed in Africa and of considerable age (39000–51400 years: Chen *et al*. 2000). Its diversity is highest in eastern and western Africa (Table 5) and it is rare in Khoisan-speaking populations. Hence it probably has its origin at latitudes immediately south of the Sahara. In Mozambique it has a reduced diversity relative to the eastern and western populations. Ten out of the fourteen Mozambique L2a HVRI haplotypes had not been observed before, and belong to two or three subclusters. The largest of these Mozambique-specific clusters in L2a (based on 16189–16290–16294–16309 on top of the L2 motif), the root sequence of which is shared with a Zimbabwean (Horai & Hayasaka, 1990), has an age of 13500±3000 years, suggesting a movement southwards in the late glacial or early Holocene. This subcluster contains another frequent haplotype (bearing 16192 in addition), which, although not observed elsewhere, is another potential founder type. If we make this assumption, the combined age falls to 6700±2100 years, a Holocene signal which could tentatively be attributed to the southern Bantu expansion.

In South African Y chromosomes from Bantu speakers, a substantial reduction in diversity is observed (Thomas *et al*. 2000): one (YAP+) haplotype, based on six microsatellite loci, together with its one-step neighbours, comprises almost half the Bantu Y chromosomes. This group of chromosomes is consistent with an expansion from a single type 3000–5000 years ago. The diversity reduction has no parallel in mtDNA, perhaps suggesting that local maternal lineages were assimilated during the expansion. Indeed, it is far from clear that the signature that we are detecting in L2a is not that of an earlier expansion, perhaps following climate change in the late-glacial, a pattern which is becoming evident in other parts of the world (Forster *et al*. 1996; Torroni *et al*. 1998; Richards *et al*. 2000), albeit in regions where the changes in the climate were rather different.

Haplogroup L3

Figure 6 shows a reduced median network of haplogroup L3 in Mozambique. All well-characterised African L3 clusters are present in Mozambique, as well as one less well-characterized group, based on 16209–16223–16292–16311 (cf. Alves-Silva *et al*. 2000), which has a distribution south of the Sahara. All samples belonged to the African-specific branches of L3. L3b comprises 4% of the sample,

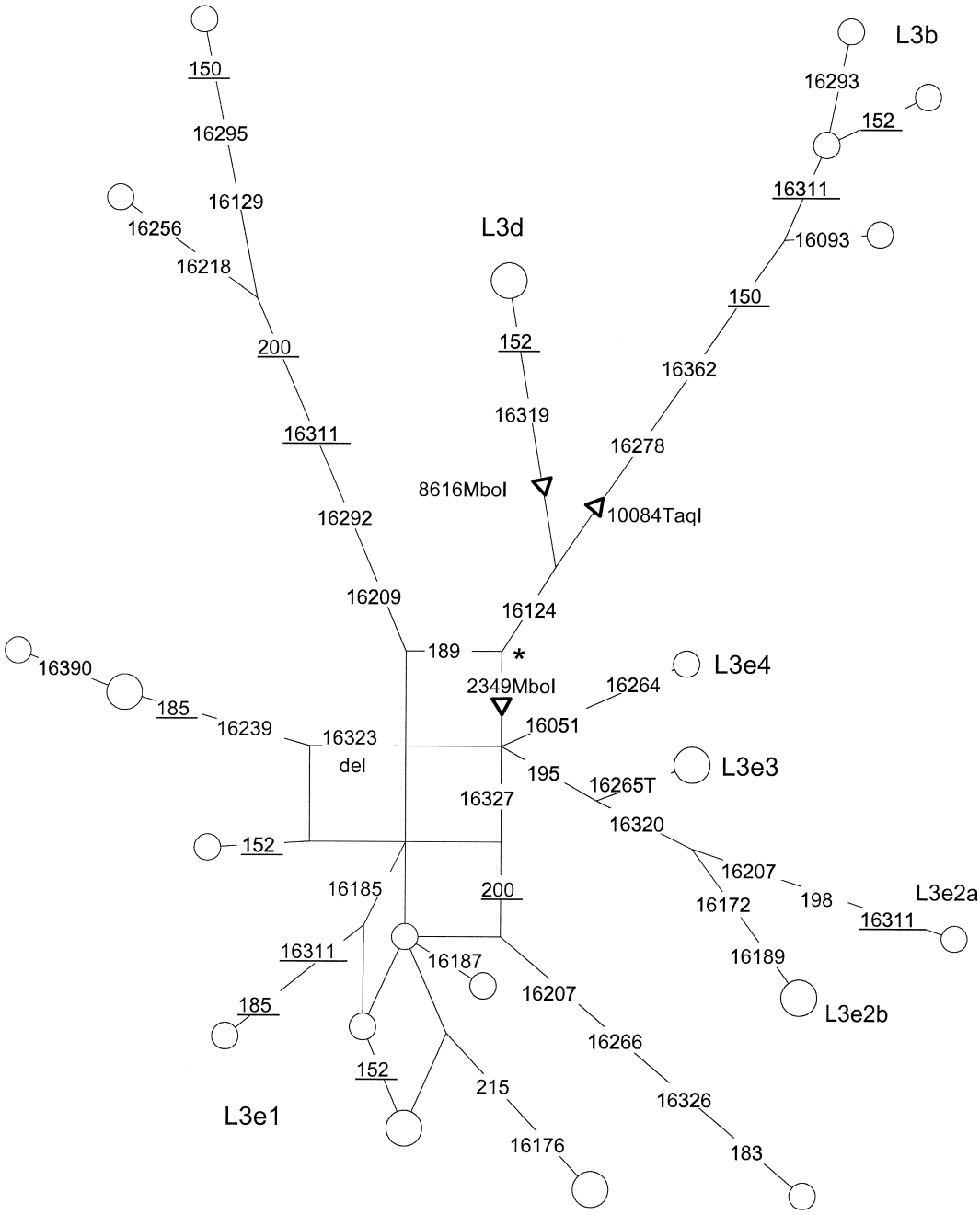

Fig. 6. The reduced median network of the 27 L3 sequences in the Mozambique sample. The circles are combined HVRI/HVRII haplotypes, the areas of which are proportional to the frequency in the sample. The smallest circles are singletons, the largest has frequency two. Also included is information on the RFLP markers assayed (the arrows indicate the direction of a site gain). Control-region mutations (shown on the branches) are transitions unless a base change is explicitly indicated. Underlining indicates resolved recurrent mutations, unresolved events being shown by reticulation. The node marked with an asterisk has the motif 16223–073–150–263. The evolution at hypervariable sites such as 150, 152 and 195 (Bandelt *et al.* 2000) is probably not accurately reconstructed. For example, there is coding-region information to suggest that L3e3 is more closely related to L3e4 than to L3e2 (Bandelt *et al.* in press).

is widespread in western Africans and has been implicated in the Bantu expansion (Watson *et al.* 1997). The single L3d haplotype presents a match with a Fulbe sequence. The full diversity of L3e is present in our sample (Bandelt *et al.* in press). We comment on two informative subclusters. L3e1a is most common in the south in both Bantu and Khoisan-speaking populations,

Table 6. *Number of matches between Mozambique* (*MOZ*), *Brazil* (*BRZ*) *and Santo Domingo* (*SD*) *sequences, and also matches for those sequences inside Africa. Information based only on HVRI between 16051 and 16362*

MOZ	BRZ[1]	SD[2]	Other	HVRI sequence	Haplo-group
1	1	—	—	093 129 148 168 172 187 188[G] 189 223 230 278 293 311 320	L1a
8	2	—	2Equatorial Guinea[3] 1MO 1KIK	129 148 168 172 187 188[G] 189 223 230 278 293 311 320	L1a
6	1	4	1Iraq[4]1TK	148 172 187 188[G] 189 223 230 311 320	L1a
1	—	1	3MAN 1SER 1SEN 1Portugal[5]	126 145 187 189 223 264 270 278 293 311	L1b
1	2	2	1SUD	129 163 187 189 209 223 278 293 294 311 360	L1c
2	1	3	2Canary[6] 1SER 1SEN 2WO 1SAH 1Syria[4] 1Equatorial Guinea[3] 1Portugal[5] 1SUD 1TU 1SO 1KIK 1FUL 1YOR	223 278 294 309	L2a
1	1	1	2KNG2	114[A] 129 213 223 278 354	L2b
1	—	4	1Galicia[7]	093 124 223 278 362	L3b
2	—	1	1SUD 2NUB 1HA	124 223 278 311 362	L3b
2	3	—	1!Kung[8]	176 223 327	L3e1*
1	1	1	—	223 327	L3e1*
3	1	—	1Dama[8]	185 223 327	L3e1a
2	3	5	2KNG2 3KWE 1WO 1Syria[4] 1HA 1FUL 2MZB 1Israel[4]	172 189 223 320	L3e2b
2	—	1	1Israel[4] 1Equatorial Guinea[3]	223 265[T]	L3e3
1	—	1	1SUD 1WO 4MAN	051 223 264	L3e4

Besides the populations referred in Table 1, the survey included other populations with the following bibliographic references: [1]Alves-Silva *et al.* (2000); [2]A. Torroni (unpublished data); [3]Mateu *et al.* (1997); [4]Richards *et al.* (2000); [5]Pereira *et al.* (2000); [6]Rando *et al.* (1999); [7]Salas *et al.* (1998); [8]Soodyall (1993).

although it has been suggested (Bandelt *et al.* in press) that its origin is further north and that it may have been carried south by the Bantu. L3e4, on the other hand, present in a single individual, was absent until now in the south and has an Atlantic western African distribution. Its presence in Mozambique presents a puzzle, since it suggests recent gene flow from the Atlantic west to southeastern Africa. A similar pattern occurs for haplogroup L1b, also present in one individual in our sample, which is concentrated in western Africa and very rare in the south (one Khwe, Chen *et al.* 2000, which differs at two positions in HVRI from the Mozambique individual).

Sequence matches

In order to investigate whether the contribution of Mozambique sequences to the mtDNA sequence pools of America was higher than in Europe, we searched for matches in a worldwide database. The African and European samples were substantial and widespread, but the Americas were represented by two populations, one from Brazil (Alves-Silva *et al.* 2000) and one from Santo Domingo, in the Caribbean (A. Torroni, unpublished data).

There was a considerable number of matches between Mozambique and American sequences from African haplogroups (Table 6), representing a total of 15 shared sequences in a total of 109 different haplotypes from African haplogroups in the American pool. Two out of these 15 matches correspond to sequences that, in Africa, have only been observed in Mozambique until now; five were not detected in western Africa (one was detected in Galicia, which probably also represents a slave introduction, three in Khoisan-speaking populations, and one in Sudan); and eight were also detected in western Africa. All the Mozambique–American matches for L3e1 were not shared elsewhere, except in Khoisan-speaking populations. With respect to those

Table 7. *Matches for sequences observed in Europe from African haplogroups to sequences in Africa, including Mozambique (MOZ). Information based only on HVRI between 16051 and 16362. African haplotypes observed in Europe but not observed in Africa: L1a (3); L1b (4); L2a (7); L3* (10); L3b (3); L3d (1); L3e (4)*

Europe	MOZ	Africa	HVRI sequence	Haplogroup
1Sardinia[9]	—	1TK 2MAN 9NUB 2SUD	129 148 168 172 187 188[G] 189 223 230 311 320	L1a
1Canary[6]	—	5FUL 1WO 1Equatorial Guinea[3]	093 126 187 189 223 264 270 278 293 311	L1b
1Portugal[5]	1	1SEN 1SER 3MAN	126 145 187 189 223 264 270 278 293 311	L1b
2Canary[6]	—	1EGY 1MO 1MA 2WO 1SER 1KWE 2Equatorial Guinea[3]	126 187 189 223 264 270 278 311	L1b
1Portugal[5]	—	1SEN 1FUL	093 189 192 223 278 294 309	L2a
1Italy[4]	10	1SO 1SEN	223 278 286 294 309	L2a
1Portugal[5]	2	1SAH 2WO 1SER	223 278 294 309	L2a
2Canary[6]		1SEN 1YOR 1FUL 1Equatorial Guinea[3] 1SUD 1TU 1SO 1KIK		
1Portugal[5]	—	1FUL 1SUD 1MO	209 223 311	L3*
1Portugal (born in Angola)[10]				
1Canary[6]				
1Portugal[5]	—	1NUB 2SUD 1MA	223	L3*
1Portugal[10]				
1Basque[4]	—	1KIK	223 311	L3*
1Switzerland[11]	—	1SUD	176 188 209 223 234 311 355	L3*
1Galicia[7]	1	—	093 124 223 278 362	L3b
1Spain[12]	—	1MO	124 223 234 278 362	L3b

[1–8] as in Table 6; [9]Di Rienzo *et al.* (1991); [10]Côrte-Real *et al.* (1996); [11]Dimo-Simonin *et al.* (2000); [12]Richards *et al.* (1996).

American sequences with no match with Mozambique, there were 25 matches out of 94 different sequences, ten restricted to western African populations.

For the European L sequence pool (Table 7), in a total of 48 different haplotypes four matches were detected with Mozambique, but three of those sequences were also detected in western African populations. Besides the matches with Mozambique, a further nine were detected, of which two were western African specific, one northern African, two eastern and four were widespread.

The comparison of the contribution to the American and to the European pools of African sequences via the Atlantic slave trade could be biased by the fact that these sequences had more ancient origins in Europe. Sub-Saharan and north African slaves are known to have been introduced during the Roman Period and also under the Muslim rule in Iberia, and the possibility of earlier, Neolithic contacts should not be discounted. Nevertheless, it is suggestive that the majority of sequences with a sub-Saharan origin within Europe are in Iberia and the Canary Islands (the first colonies of the Iberian kingdoms), which were most extensively involved in the slave trade.

There remains a large number of sequences from African haplogroups sampled in the Americas and Europe for which no match can be found in the current African database. This may be due in part to the fact that the main regions from where slaves were taken, such as Angola and the Slave Coast (Thomas, 1998), remain uncharacterized.

Conclusions

The phylogeographic analysis of the Mozambique sequences in this study has revealed distinct components from the north and the

south. An influence from Khoisan-speaking populations was detected as judged by the considerable proportion of distinct L1d sequences, a possible relict of the populations that inhabited this region before the arrival of the Bantu speakers. The Bantu expansion, although originating in a single western core, proceeded in two directions, western and eastern, both towards the south, although only the second reached the very south of Africa. Whether the Bantu mtDNA pool was or was not different in the west and the east remains to be clarified. Comparison with, for instance, Angolans, where some Khoisan-speaking groups are still present, would be essential in order to evaluate Bantu and Khoisan influences in both African coasts. Nowadays in Mozambique, there are no Khoisan-speaking ethnic groups. Although there has been a linguistic replacement, it is unlikely to have been a complete population replacement, as evidenced by the L1d types.

As possible remnants of the Bantu expansion through east towards south Africa, we detected all the haplogroups that have been implicated in this expansion, that is L3b, L3e1a and a subset of L1a sequences. A tentative dating of some L2a sequences, the most frequent haplogroup in the Mozambique sample, by postulating two founder types, as suggested by their low diversity and star-like phylogenies, displayed an age range overlapping the Bantu expansion, although an earlier arrival of these types cannot be excluded. Recent gene flow from Atlantic Africa seems the most probable explanation for the detection of one L1b and one L3e4 sequence in Mozambique.

With respect to the eastern African slave input to America and to Europe, the higher proportion of matches between sequences from Mozambique and the Americas compared to that between Mozambique and Europe, is in accordance with the historical documentation (Thomas, 1998) of a differential slave trade, with eastern African slaves more likely to be taken to the Americas. This is particularly striking since other documented factors would have tended to weaken this signal. Firstly, the female/male proportion of the slaves taken to Europe was much higher than for those taken to America, and secondly, slave reproduction (particularly from female slaves and white owners) was stimulated in Europe (especially after the ban on the importation of slaves after the middle of the eighteenth century), but has been repressed in America (Thomas, 1998).

W e thank Martin Richards and an anonymous referee for suggesting improvements to the manuscript. This work was supported by the following grants: a PhD grant (PRAXIS BD/13632/97) financed by Fundação para a Ciência e a Tecnologia to LP and to IPATIMUP through Programa Operacional Ciência, Tecnologia e Inovação (POCTI), Quadro Communitário de Apoio III; a Wellcome Trust Career Development Fellowship to VM; by the 'Istituto Pasteur Fondazione Cenci Bolognetti', Università di Roma 'La Sapienza' (to R.S.), Grandi Progetti Ateneo, Università di Roma 'La Sapienza' (to R.S.), Consiglio Nazionale delle Ricerche (99.02620.CT04) (to A.T.), Telethon-Italy E.0890 (to A.T.), Fondo d'Ateneo 2001 dell'Università di Pavia (to A.T.), the Italian Ministry of the University, Progetti Ricerca Interesse Nazionale 1999 and 2001 (to R.S and A.T.). The Mozambique samples were kindly provided by Dr. Albertino Damasceno and Dr. Benilde Soares of the Eduardo Mondlane University (Maputo).

REFERENCES

Alves-Silva, J., Santos, M. D. S., Guimarães, P. E. M., Ferreira, A. C. S., Bandelt, H.-J., Pena, S. D. J. & Prado, V. F. (2000). The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* **67**, 444–461.

Anderson, S., Bankier, A. T., Barrell, B. G., De Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I., Nierlich, D., Roc, B., Sanger, F., Schreier, P., Smith, A., Staden, A. & Young, I. (1981). Sequence and organisation of the human mitochondrial genome. *Nature* **290**, 457–465.

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. & Forster, P. (1997). The myth of bumpy hunter-gatherer mismatch distributions. *Am. J. Hum. Genet.* **61**, 980–983.

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. Phylogenet. Evol.* **16**, 8–28.

Bandelt, H.-J., Alves-Silva, J., Guimarães, P. E. M., Santos, M. S., 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. Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann. Hum. Genet.* **65**, in press.

Bendall, K. E. & Sykes, B. C. (1995). Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am. J. Hum. Genet.* **57**, 248–256.

Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J. & Allen, R. C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* **48**, 137–144.

Cann, R. L., Stoneking, M. & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.

Chen, Y.-S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S. & Wallace, D. C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57**, 133–149.

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., Macaulay, V., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, A., Bertranpetit, J. & Sykes, B. (1996). Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.* **60**, 331–350.

Di Rienzo, A. & Wilson, A. C. (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **88**, 1597–1601.

Dimo-Simonin, N., Grange, F., Taroni, F., Brandt-Casadevall, C. & Mangin, P. (2000). Forensic evaluation of mtDNA in a population from south west Switzerland. *Int. J. Legal Med.* **113**, 89–97.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package). Distributed by the author, Department of Genetics, University of Washington.

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.

Horai, S. & Hayasaka, K. (1990). Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* **46**, 828–842. .

Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., Harihara, S., Park, K. S., Omoto, K. & Pan, I. H. (1996). mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* **59**, 579–590.

Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713.

Kivisild, T., Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W. S., Dixon, M. E., Papiha, S. S., Mastana, S. S., Mir, M. R., Ferak, V. & Villems, R. (1999). Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Current Biology* **9**, 1331–1334.

Krings, M., Salem, A.H., Bauer, K., Geisert, H., Malek, A., 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.

Lareu, M. V., Phillips, C. P., Carracedo, A., Lincoln, A. J., Syndercombe-court, D. & Thomson, J. A. (1994). Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats; UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic Sci. Int.* **66**, 41–52.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-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.

Mateu, E., Comas, D., Calafell, F., Pérez-Lezaun, A., Abade, A. & Bertranpetit, J. (1997). A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann. Hum. Genet.* **61**, 507–518.

Newman, J. L. (1995).*The peopling of Africa: a geographic interpretation.*Yale University Press.

Pereira, L., Prata, M. J. & Amorim, A. (2000). mtDNA diversity in Portugal: not a genetic edge of European variation. *Ann. Hum. Genet.* **64**, 491–506.

Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H.-J. (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.

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., Côrte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J. & Sykes, B. (1996). Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**, 185–203.

Richards, M. & Macaulay, V. (2000). *Genetic data and the colonization of Europe: genealogies and founders.* In Renfrew, C. & Boyle, K. Archaeogenetics: DNA and the Population Prehistory of Europe. Cambridge: McDonald Institute for Archaeological Research, pp. 139–151.

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., Calí, 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., Scozzari, R, Torroni, A., & Bandelt, H.-J. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276.

Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999). Genetic evidence for an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.* **23**, 437–441.

Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.-J. & Nørby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* **67**, 718–726.

Salas, A., Comas, D., Lareu, M. V., Bertranpetit, J. & Carracedo, A. (1998). mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur. J. Hum. Genet.* **6**, 365–375.

Schneider, S., Roessli, D. & Excoffier, L. (2000). Arlequin ver. 2.0: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Soodyall, H. (1993). Mitochondrial DNA polymorphisms in Southern African populations. PhD thesis, University of the Witwatersrand, Johannesburg.

Soodyall, H., Vigilant, L., Hill, A. V., Stoneking, M. & Jenkins, T. (1996). mtDNA control-region sequence variation suggests multiple origins of an 'Asian-specific' 9-bp deletion in sub-Saharan Africans. *Am. J. Hum. Genet.* **58**, 595–608.

Thomas, H. (1998). *The slave trade – the history of the Atlantic slave trade 1440–1870*. London: Macmillan Publishers Ltd.

Thomas, M. G., Parfitt, T., Weiss, D. A., Skorecki, K., Wilson, J. F., le Roux, M., Bradman, N. & Goldstein, D. B. (2000). Y chromosome traveling south: the Cohen modal haplotype and the origins of the Lemba – the 'black Jews of southern Africa'. *Am. J. Hum. Genet.* **66**, 674–686.

Torroni, A., Schurr, T. G., Yang, C.-C., Szathmary, E. J. E., Williams, R. C., Schanfield, M. S., Troup, G. A., Knowler, W. C., Lawrence, D. N. & Weiss, K. M. (1992). Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* **130**, 153–162.

Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savantaus, M.-L., Bonné-Tamir, B. & Scozzari, R. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* **62**, 1137–1152.

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.

Appendix 1. *Control region sequences in Mozambique*

Variant positions from the CRS are shown between 16017 and 16390 in HVRI (minus 16000) and 73–340 in HVRII. The substitution indicated in italics was observed to be heteroplasmic. Substitutions are transitions unless the base change or a deletion is explicitly indicated. Insertions of one and two cytosines are shown by appending '.1' and '.2', respectively. Also shown are the results of the partial RFLP typing. A = 3592*Hpa*I; B = 10084*Taq*I; C = 8616*Mbo*I; D = 2349*Mbo*I. + indicates the presence of the restriction site, − the absence, n/a = not available. The haplogroup assignment is discussed in the text.

Freq.	HVRI sequence	HVRII sequence	RFLP				Haplo-group
			A	B	C	D	
1	93 129 148 168 172 187 188$^{C/G}$ 189 223 230 278 293 311 320	93 95$^{A/C}$ 185 189 236 247 263 311.1	n/a	n/a	n/a	n/a	L1a
5	129 148 168 172 187 188$^{C/G}$ 189 223 230 278 293 311 320	93 95$^{A/C}$ 185 189 236 247 263 303.1 311.1	n/a	n/a	n/a	n/a	L1a
1	129 148 168 172 187 188$^{C/A}$ 189 223 230 278 293 311 320	93 95$^{A/C}$ 185 189 236 247 263 303.1 311.1	n/a	n/a	n/a	n/a	L1a
2	129 148 168 172 187 188$^{C/G}$ 189 223 230 278 293 311 320	93 95$^{A/C}$ 185 189 236 247 263 311.1	n/a	n/a	n/a	n/a	L1a
1	129 148 168 172 187 188$^{C/G}$ 189 223 230 278 293 311 320	93 95$^{A/C}$ 185 189 236 247 263 303.2 311.1	n/a	n/a	n/a	n/a	L1a
1	148 172 187 188$^{C/G}$ 189 223 230 311 320	93 150 152 189 204 207 236 247 263 311.1	n/a	n/a	n/a	n/a	L1a
1	148 172 187 188$^{C/G}$ 189 223 230 311 320	93 152 189 199 204 207 236 247 263 311.1	n/a	n/a	n/a	n/a	L1a
3	148 172 187 188$^{C/G}$ 189 223 230 311 320	93 152 189 204 207 236 247 263 311.1	n/a	n/a	n/a	n/a	L1a
1	148 172 187 188$^{C/G}$ 189 223 230 311 320	93 152 189 236 247 263 311.1	n/a	n/a	n/a	n/a	L1a
1	126 145 187 189 223 264 270 278 293 311	73 152 182 185$^{G/T}$ 195 247 263 311.1 357	n/a	n/a	n/a	n/a	L1b
1	17 129 163 187 189 209 223 278 293 294 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 247 263 311.1 316	n/a	n/a	n/a	n/a	L1c
1	17 129 187 189 223 278 293 294 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 247 263 303.1 311.1 316	n/a	n/a	n/a	n/a	L1c
1	71 129 145 187 189 213 223 234 265$^{A/C}$ 278 286$^{C/G}$ 294 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 297 311.1 316	n/a	n/a	n/a	n/a	L1c
1	71 145 187 189 213 223 234 265$^{A/C}$ 278 286$^{C/G}$ 294 311 360	73 93 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 297 303.1 311.1 316	n/a	n/a	n/a	n/a	L1c
1	129 183$^{A/C}$ 189 215 223 278 294 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 247 263 303.1 311.1 316	n/a	n/a	n/a	n/a	L1c
1	129 145 187 189 212 223 230 243 311 390	73 146 152 195 198 247 311.1	n/a	n/a	n/a	n/a	L1d
1	129 187 189 212 223 230 243 291 311	73 146 152 188 195 247 295 311.1	n/a	n/a	n/a	n/a	L1d
1	129 187 189 212 223 230 243 291 311	73 146 152 188 195 228 247 311.1	n/a	n/a	n/a	n/a	L1d
2	129 187 189 223 230 243 311	73 146 152 195 247^{delG} 294$^{T/A}$ 311.1	n/a	n/a	n/a	n/a	L1d
1	129 187 189 223 230 243 311 390	73 146 152 195 198 247 311.1	n/a	n/a	n/a	n/a	L1d
1	129 187 189 223 239 243 294 311	73 146 152 195 247 311.1	n/a	n/a	n/a	n/a	L1d
1	187 189 223 230 234 243 294$^{C/G}$ 311	73 146 152 195 247 311.1	n/a	n/a	n/a	n/a	L1d
1	129 148 166 183^{delA} 187 189 192 223 278 311 355 362	73 152 182 247 263 311.1	n/a	n/a	n/a	n/a	L1e
1	129 148 166 183^{delA} 187 189 192 223 278 311 355 362 390	73 152 182 195 247 263 311.1	n/a	n/a	n/a	n/a	L1e
1	223 234 278 294 295 390	73 146 152 195 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a

Appendix 1 (*cont.*)

Freq.	HVRI sequence	HVRII sequence	RFLP				Haplo-group
			A	B	C	D	
1	51 182$^{A/C}$ 183$^{A/C}$ 189 192 223 278 290 294 304 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	79 223 278 286 294 309 390	73 146 152 195 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a
1	102 182$^{A/C}$ 183$^{A/C}$ 189 223 278 290 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	129 182$^{A/C}$ 183$^{A/C}$ 189 223 278 290 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	131 189 223 234 278 294 309 390	73 146 152 263 303.2 311.1	n/a	n/a	n/a	n/a	L2a
1	168 182$^{A/C}$ 183$^{A/C}$ 189 223 278 290 294 309 390	73 146 152 195 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a
3	182$^{A/C}$ 183$^{A/C}$ 189 192 223 278 290 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	182$^{A/C}$ 183$^{A/C}$ 189 192 223 278 290 294 309 390	73 146 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	182$^{A/C}$ 183$^{A/C}$ 189 223 278 290 294 309 390	73 146 152 195 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a
2	182$^{A/C}$ 183$^{A/C}$ 189 223 224 278 290 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
11	182$^{A/C}$ 183$^{A/C}$ 189 223 278 290 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	182$^{A/C}$ 183$^{A/C}$ 189 223 278 290 294 309 390	73 146 152 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a
4	182$^{A/C}$ 183$^{A/C}$ 189 192 223 278 290 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	182$^{A/C}$ 183$^{A/C}$ 189 192 223 278 290 294 309 390	73 146 152 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a
1	182$^{A/C}$ 183$^{A/C}$ 189 223 278 290 294 296 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
5	223 278 286 294 309 390	73 146 152 195 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a
2	223 278 286 294 309 390	73 146 152 195 263 303.2 311.1	n/a	n/a	n/a	n/a	L2a
1	223 278 286 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	223 278 286 294 309 390	73 146 152 263 303.2 311.1	n/a	n/a	n/a	n/a	L2a
1	223 278 286 294 309 390	73 146 195 263 311.1	n/a	n/a	n/a	n/a	L2a
3	223 278 286 294 390	73 146 152 195 263 303.1 311.1	n/a	n/a	n/a	n/a	L2a
1	223 278 294 309 390	73 146 152 195 263 303.2 311.1	n/a	n/a	n/a	n/a	L2a
1	223 278 294 309 390	73 146 152 195 263 311.1	n/a	n/a	n/a	n/a	L2a
1	114$^{C/A}$ 129 213 223 274 278 390	73 146 150 152 182 183 195 198 204 263 303.1 311.1	n/a	n/a	n/a	n/a	L2b

Appendix 1 (*cont.*)

Freq.	HVRI sequence	HVRII sequence	RFLP				Haplo-group
			A	B	C	D	
1	114$^{C/A}$ 129 213 223 278 354 390	73 146 150 152 182 195 198 204 263 311.1	n/a	n/a	n/a	n/a	L2b
1	134 223 264 278 390	73 93 146 150 152 182 195 198 263 311.1 325	n/a	n/a	n/a	n/a	L2c
1	129 209 223 292 295 311	73 189 *200* 263 303.1 311.1	−	−	+	−	L3*
1	209 218 223 256 292 311	73 150 189 200 263 311.1	−	−	+	−	L3*
1	93 124 223 278 362	73 263 311.1	−	+	+	−	L3b
1	124 223 278 293 311 362	73 263 303.1 311.1	−	+	+	−	L3b
1	124 223 278 311 362	73 152 263 311.1	−	+	+	−	L3b
1	124 223 278 311 362	73 263 311.1	−	+	+	−	L3b
2	124 223 319	73 150 152 263 303.1 311.1	−	−	−	−	L3d
2	223 239 323^{delT}	73 150 185 189 263 303.1 311.1	−	−	+	+	L3e1*?
1	223 239 323^{delT} 390	73 150 185 189 263 303.1 311.1	−	−	+	+	L3e1*?
1	223 323^{delT} 327	73 150 152 189 263 303.1 311.1	−	−	+	+	L3e1*
2	176 223 327	73 150 152 189 200 215 263 311.1	−	−	+	+	L3e1*
1	187 223 327	73 150 189 200 263 311.1	−	−	+	+	L3e1*
1	207 223 266 326 327	73 150 183 200 263 303.1 311.1	−	−	+	+	L3e1*
1	223 327	73 150 189 200 263 311.1	−	−	+	+	L3e1*
1	185 223 311 327	73 150 185 189 263 311.1	−	−	+	+	L3e1a
2	185 223 327	73 150 152 189 200 263 311.1	−	−	+	+	L3e1a
1	185 223 327	73 150 189 200 263 303.1 311.1	−	−	+	+	L3e1a
1	207 223 311 320	73 150 195 198 263 311.1	−	−	+	+	L3e2a
2	172 183$^{A/C}$ 189 223 320	73 150 195 263 311.1	−	−	+	+	L3e2b
2	223 265$^{A/T}$	73 150 195 263 311.1	−	−	+	+	L3e3
1	51 223 264	73 150 263 311.1	−	−	+	+	L3e4

Total: 109