Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 43 (2007) 635–644
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa

Chiara Batini [a], Valentina Coia [a], Cinzia Battaggia [a], Jorge Rocha [b], Maya Metni Pilkington [c], Gabriella Spedini [a,e], David Comas [d], Giovanni Destro-Bisol [a,e,*], Francesc Calafell [d]

[a] *Dipartimento di Biologia Animale e dell'Uomo, Università "La Sapienza", Roma, Italy*
[b] *Instituto de Patologia e Imunologia Molecular (IPATIMUP), Universidade do Porto, Porto, Portugal*
[c] *Department of Anthropology, University of Arizona, Tucson, USA*
[d] *Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat "Pompeu Fabra", Barcelona, Spain*
[e] *Istituto Italiano di Antropologia, Roma, Italy*

Received 28 April 2006; revised 3 August 2006; accepted 8 September 2006
Available online 5 October 2006

Abstract

Interindividual variation of human mitochondrial DNA has been extensively studied over the last two decades, and its usefulness for reconstructing evolutionary relationships of extant populations has been proved. However, some mitochondrial lineages still need to be studied using a combination of larger and tailored datasets and increased level of resolution in order to shed light on their origin and on the processes underlying their present distribution. In this study, we analyze the phylogeny of the L1c haplogroup of human mitochondrial DNA using sequence data from hypervariable regions 1 and 2 obtained from 455 individuals (extracted from a total sampling of 2542 individuals) belonging to sub-Saharan African and African-American populations. We propose a substantial revision of L1c phylogeny, by introducing one new sub-haplogroup (L1c4), two new L1c1 clades (L1c1b and L1c1c), and by reassigning the previous L1c1a1 sequences to a clade which we termed L1c5. The new phylogeny encompasses distinct lineages with different evolutionary histories. In fact, based on population frequency, internal variation and mismatch distribution, we propose that L1c1b, L1c1c and L1c2 originated in Bantu ancestors, whereas L1c1a, L1c4 and L1c5 evolved among Western Pygmies. The population structure of L1c is not comparable to any known mitochondrial or, even, Y-chromosomal haplogroup, and challenges the current view that most of mtDNA variation in Pygmies might reflect admixture with Bantu or a persistence of plesiomorphic characters. In fact, the unique feature of the L1c is that it retains a signature of a phase common to the ancestors of the Bantu and Western Pygmies, while encompassing some specific sub-clades which can indicate their divergence. This allowed us to attempt a phylogenetically based assessment of the evolutionary relationships between the two groups. Taking into consideration estimates of the time to the most recent common ancestor of L1c and its clades together with archaeological and paleoclimatological evidence, we propose that the ancestors of Bantu and Western Pygmies separated between 60 and 30 kya.

Keywords: Mitochondrial DNA; HVR-I; HVR-II; Western Pygmies; Bantu; Middle Stone Age

1. Introduction

Human mitochondrial DNA (mtDNA) is a circular double-stranded molecule, consisting of 16569 base pairs (Anderson et al., 1981). Since the early 1980s it has been widely used for human population studies, because of its high copy number, maternal inheritance, apparent lack of recombination and high mutation rate (see Pakendorf and Stoneking, 2005, for a review). In the last twenty-five years, numerous data regarding mitochondrial DNA variation in human populations have been accumulated, with an increasing level of resolution (see Richards and Macaulay, 2001, for a review).

One of the pioneer studies of mtDNA variation in human populations produced a tree that showed a deep split between sub-Saharan Africans and non-Africans with

* Corresponding author. Fax: +39 06 4991 2771.
E-mail address: destrobisol@uniroma1.it (G. Destro-Bisol).

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.09.014

a coalescence which dates back to 200 thousand years ago (kya; Cann et al., 1987). This was interpreted as evidence supporting a recent origin of modern humans in Africa. The Cann et al. study (1987) paved the way for further investigations into human populations, of which sub-Saharan African populations were considered to be of particular interest and importance because of their essential role in any genetic test of the hypotheses concerning the emergence of modern humans (Harpending et al., 1993). It was observed early on that most sub-Saharan mtDNAs (from 70% to 100%, depending on the population considered) present a specific HpaI restriction site at position 3592 (Torroni et al., 1994). These haplotypes were subsequently assigned to a lineage which was conventionally termed L (Chen et al., 1995; see Salas et al., 2002, for a review) and which contains several superhaplogroups (L0, L1, L2, and L3; Salas et al., 2004), with relative haplogroups.

The L1c haplogroup of human mitochondrial DNA was first defined by Rando et al. (1998) on the basis of transitions in the hypervariable region 1 (HVR-1) (at np 16129–16187–16189–16223–16278–16294–16311–16360), transversions in the hypervariable region 2 (HVR-2) (at np 186 and 189), and gain of a TaqI (np 9070) and RsaI (np 12810) restriction site in the coding region. The mtDNA studies carried out so far make it possible to give a primary indication of L1c distribution, showing that it occurs at the highest frequencies in Central Africa (2–96%), whereas it is less common in North (4–7%), West (2–10%) and East Africa (1–5%) (see references of Table 1; Watson et al., 1997; Rando et al., 1998; Krings et al., 1999; Alves-Silva et al., 2000; Green et al., 2000; Salas et al., 2002; Fadhlaoui-Zid et al., 2004). L1c has also been found in American populations of African descent, with different frequencies in North (11%), Central (2–8%) and South America (19%) (Table 1; Alves-Silva et al., 2000).

The origin of L1c and the processes leading to its present distribution are still a matter of debate, due both to lack of data concerning some crucial areas and to the low level of resolution used in most studies. The first study in which L1c was found to occur at high frequencies (19%) was

Table 1
Population database for hypervariable regions 1 and 2 of the mitochondrial DNA used in this study

Population	Country	L1c freq	*N*	Source	
				HVR-1	HVR-2
West Africa					
Mandenka	Senegal	0.042	5	Graven et al. (1995)	Graven et al., 1995
Central Africa					
Babinga	PRC	0.863	38	Present study	Present study
Baka	Cameroon	0.840	42	Present study	Present study
Bakaka	Cameroon	0.140	7	Coia et al. (2005)	Present study
Bakola	Cameroon	1.000	49	Present study	Present study
Bamileke	Cameroon	0.060	3	Coia et al. (2005)	Present study
Bassa	Cameroon	0.240	10	Coia et al. (2005)	Present study
Bateke	PRC	0.140	7	Present study	Present study
Biaka	CAR	0.764	13	Vigilant et al. (1991)	Vigilant et al. (1991)
Daba	Cameroon	0.100	2	Coia et al. (2005)	Present study
Ewondo	Cameroon	0.142	7	Destro-Bisol et al. (2004a)	Present study
Fali	Cameroon	0.020	1	Coia et al. (2005)	Present study
Fulbe Cameroon	Cameroon	0.060	1	Coia et al. (2005)	Present study
Mandara	Cameroon	0.054	2	Coia et al. (2005)	Present study
Mbenzele	CAR	0.958	47	Destro-Bisol et al. (2004a)	Present study
Ngoumba	Cameroon	0.136	6	Present study	Present study
Sanga	CAR	0.400	8	Present study	Present study
Tali	Cameroon	0.050	1	Coia et al. (2005)	Present study
Uldeme	Cameroon	0.040	1	Coia et al. (2005)	Present study
East Africa					
Mozambique	Mozambique	0.045	5	Pereira et al. (2001)	Pereira et al. (2001)
South-West Africa					
Angola	Angola	0.159	7	Plaza et al. (2004)	Plaza et al. (2004)
Cabinda	Angola	0.245	27	Beleza et al. (2005)	Beleza et al. (2005)
Insular Central Africa					
São Tomé	São Tomé	0.209	20	Unpublished data, J. Rocha	Unpublished data, J. Rocha
São Tomé	São Tomé	0.193	16	Trovoada et al. (2004)	Trovoada et al. (2004)
America					
F.B.I, database	USA	0.108	124	Monson et al. (2002)	Monson et al. (2002)
Choco	Colombia	0.081	4	Salas et al. (2005)	Salas et al. (2005)
Garifuna	Honduras	0.045	2	Salas et al. (2005)	Salas et al. (2005)

Note: *N* = number of individuals L1c analyzed in the study.

conducted by Alves-Silva et al. (2000) on a Brazilian population of partial African ancestry. The authors proposed that Angola could be a L1c *reservoir* since this area is the major source of African slaves brought to Brazil. This hypothesis has been subsequently supported by recent surveys carried out in Bantu populations from Angola by Plaza et al. (2004) and Beleza et al. (2005), who observed a high frequency of L1c (16–24%). However, the above studies also observed that Angolan L1c sequences lie at the tip of the phylogeny of the haplogroup, far from the root sequence, which is in contrast with a local origin of this haplogroup. Considering this, Plaza et al. (2004) argued that a study of populations inhabiting the area between Cameroon and the Popular Republic of Congo is needed to gain further insights into the L1c origin. Salas et al. (2002) suggested that this haplogroup could be an indigenous lineage assimilated in the forest zone by Western Bantu flow. The evidence that L1c reaches its highest frequencies in Mbenzele (96%), Western Pygmies from the Central African Republic (Destro-Bisol et al., 2004a) supports this view, and indicates that a more systematic investigation of populations inhabiting Central Africa before the Bantu expansion is an important step to better understand L1c evolution.

In the present study, we analyze sequences belonging to 18 populations from Cameroon, the Central African Republic (CAR), the Popular Republic of Congo (PRC) and São Tomé, which add to the available L1c dataset for hypervariable regions 1 and 2. This implementation makes the dataset particularly suitable for a study of L1c evolution. In fact, it includes three additional Western Pygmy populations (Baka and Bakola from Cameroon, and Babinga from the Popular Republic of Congo), and one Bantu population from the Popular Republic of Congo (Bateke), a region which is very close to the supposed area of origin of L1c but which has yet to be analyzed for human mtDNA variation. We propose a substantial revision of L1c phylogeny, which is shown to encompass distinct lineages with different evolutionary histories, and discuss the implications of our findings for the evolutionary relationships between Western Pygmies and Bantu.

2. Materials and methods

2.1. Populations analyzed and the L1c database

Our database consists of a total of 455 individuals bearing haplogroup L1c, who have been sequenced for HVR-1 and HVR-2 (Table 1). Data relative to 203 individuals were obtained from previous studies, where 1742 individuals were examined (Vigilant et al., 1991; Graven et al., 1995; Pereira et al., 2001; Monson et al., 2002; Trovoada et al., 2004; Plaza et al., 2004; Beleza et al., 2005; Salas et al., 2005). The remaining 252 subjects (obtained from 18 populations) were selected from a total of 800 individuals in the course of this study: 150 were sequenced *ex novo* and 102 were sequenced only for HVR-2, since their HVR-1 data were already available (Destro-Bisol et al., 2004a; Coia et al., 2005). The dataset comprises six populations (Babinga, and Bateke from the Popular Republic of Congo; Baka, Bakola, and Ngoumba from Cameroon; Sanga from the Central African Republic) not yet analyzed for mtDNA variation. All HVR-1 sequences of the populations mentioned above (a total of 267 individuals) are shown in the supplementary material (Table S1). A table with the entire L1c haplotype dataset used in the present work is also provided as supplementary material (Table S2).

2.2. Laboratory analyses

Specimens collected in K_3EDTA were maintained at 4 °C until their arrival at the Laboratory of Anthropology of the University of Rome, La Sapienza. Genomic DNA was extracted from blood following a standard phenol–chloroform protocol (Gill et al., 1985). It was then quantified by direct comparison in agarose minigels.

Sequencing of the hypervariable region 1 (HVR-1) was carried out according to Vigilant et al. (1989), with minor modifications. Hypervariable region 1 was amplified using primers L15996 and H16401 and then purified using High Pure PCR Product Purification Kit (Roche). The sequence reaction was performed with the AmpliCycle Sequencing kit (Applied Biosystems) and the amplified products were analyzed in a polyacrylamide denaturing gel using a semi-automated DNA Sequencer (A.L.F. Express, Pharmacia Biotech, Uppsala, Sweden). For each sample, the sequence between positions 16040 and 16370 was determined.

For those samples that were classified into the L1c haplogroup, the hypervariable region 2 (HVR-2) was amplified using primers L29 and H408 and then purified. The sequence reaction was performed with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), and the sequence products were run on an ABI PRISM 3100 sequencer (Applied Biosystems). For each sample the sequence between positions 070 and 340 was determined.

2.3. Data analyses

A median joining network (Bandelt et al., 1995) of the human L1c haplogroup was drawn with the aid of Network 3.0 package (Fluxus Engineering Web site). Length variations at positions 16184–16193 in the HVR1 and 303–315 in the HVR2 were not considered, because of their high mutation rate which may be a source of error in phylogenetic reconstructions (Santos et al., 2005). We adopted two criteria to classify sequences into the L1c haplogroup: first we considered all the sequences assigned to L1c, taking into consideration the HVR-1 only, on the basis of the phylogeny of Salas et al. (2002); after HVR-2 sequencing, we selected all those sequences that presented the two diagnostic transversions in this region (186A and 189C; Rando et al., 1998). The network was further resolved using additional information concerning the mutability at different positions (Meyer et al., 1999). Each

nucleotide position was given a weight which was inversely proportional to its relative mutation rate (Meyer et al., 1999), so that a site with a relative mutation rate μ would be given a weight $10/\mu$. Thus, more stable sites carried larger weights in resolving reticulations.

The control of the network for phantom mutations was carried out building a network without considering the speedy mutations and detecting the presence of hypercubes (Bandelt et al., 2002). The dataset produced simple reticulations, indicating the absence of phantom mutations (Figure S1).

Haplotype diversity (HD), mean number of pairwise differences (MNPD), mismatch distributions and raggedness values (r) were calculated with the aid of Arlequin 2.0 software (Schneider et al., 2000).

The time to the most recent common ancestor (TMRCA) of L1c and its sub-clades in the phylogeny was estimated as described by Forster et al. (1996) and Saillard et al. (2000).

3. Results

The final L1c dataset contains HVR-1 and HVR-2 sequences from 455 different subjects. The haplotype diversity and mean number of pairwise differences are 0.979 ± 0.003 and 11.269 ± 5.124, respectively. The values for these measures of diversity obtained using the HVR-1 only (0.966 ± 0.004; 6.96) are very close to that obtained by Salas et al. (2002) (0.968 ± 0.010; 5.53). The TMRCA for the entire haplogroup is $91{,}250 \pm 12{,}700$ ya, considerably older than the previous estimate ($59{,}650 \pm 11{,}800$ ya; Salas et al., 2002), based on a smaller dataset (111 individuals).

The simultaneous use of both hypervariable regions produces a new L1c phylogeny and, consequently, makes a revision of the nomenclature necessary. In order to maintain a structure which was congruent with that of Salas et al. (2002), we continued to use the nucleotide status at the 16293 position as the criterion to separate L1c1 from other L1c sub-clades. The structure of our L1c phylogenetic

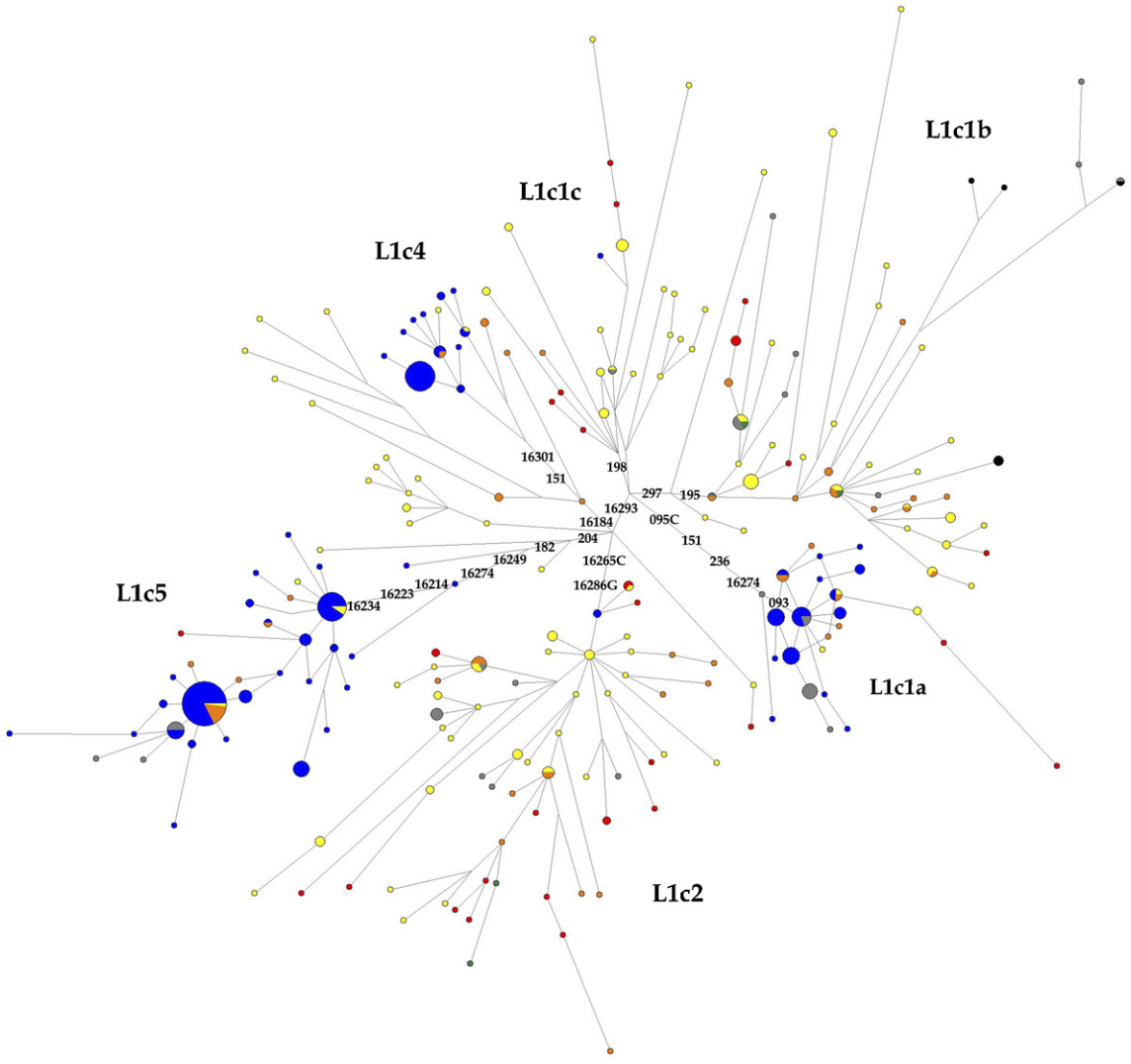

Fig. 1. Median Joining Network of L1c human mitochondrial DNA haplogroup. The diagnostic positions for the identification of the sub-clades of L1c and the new nomenclature are shown; a single letter suffix indicates a transversion. Blue, Western Pygmies; Orange, Central Africa; Red, South West Africa; Grey, Insular Central Africa; Green, East Africa; Black, West Africa; Yellow, African American.

Table 2
Definition of L1c sub-haplogroups, and correspondence with the nomenclature in Salas et al. (2002)

Presentwork	Salas et al. (2002)	Diagnostic positions
L1c (root type)	L1c (root type)	16129–16187–16189–16223–16278–16294–16311–16360 (073–151–152–182–186A–189C–195–247–263–297–316)
L1c1	L1c1	16293
L1c1a	L1c1a	16274–16293 (093–095C–**151**–236)
L1c1b	—	16293 (**195**–**297**)
L1c1c	—	16293 (198)
L1c2	L1c2	16265C–16286G
L1c4	—	16184–16301 (**151**)
L1c5	L1c1a1	16214–**16223**–16234–16249–16274 (**182**–204)

Note: The nucleotide positions listed are transitions with respect to L1c root type, unless otherwise indicated. The positions in bold indicate a loss of the transition with respect to the L1c root type. Positions in parentheses were not included in the definition by Salas et al. (2002) since they correspond to HVR-2. Note that L1c3 in Salas et al. (2002) is included within L1c1b.

tree is shown in Fig. 1. It is characterized by the presence of six clades, five of which are well defined (L1c1a, L1c1b, L1c2, L1c4 and L1c5) and one further haplogroup (L1c1c) defined on the basis of a fast evolving site in the HVR-2 (198). It would be useful to further analyze sub-clades whose identification relies on fast evolving sites, such as 16293 in HVR1 or 198 in HVR-2, by using complete-sequence data. Unfortunately, the complete-sequence data available at the moment, including both hypervariable and coding regions, do not allow a robust reconstruction of L1c phylogeny to be made (Ingman and Gyllensten, 2006). However, the large number of sequences considered makes the tree relatively robust and the consequent inferences worth of discussion. The comparison between the new nomenclature and the one used by Salas et al. (2002) is shown in Table 2. The L1c distribution in Africa and America according to this new nomenclature is presented in the Table 3.

Essentially, our network differs from that of Salas et al. (2002) in three important aspects. Firstly, we introduce two new L1c1 clades (in addition to L1c1a): L1c1b, which contains sequences previously assigned to either L1c1 or L1c3, and L1c1c, mostly composed of African American sequences. Secondly, the sequences previously assigned to

Table 3
Distribution of L1c and its sub-clades in the populations analyzed in this work

	L1c*	L1c1a	L1c1b	L1c1c	L1c2	L1c4	L1c5	L1c(tot)
West Africa								
Mandenka			0.042					0.042
Central Africa								
Babinga		0.159				0.704		0.863
Baka	0.040	0.080		0.020	0.040	0.020	0.640	0.840
Bakaka					0.100		0.040	0.140
Bakola		0.265					0.735	1.000
Bamileke			0.020		0. 040			0.060
Bassa		0.044	0.044	0.022	0.130			0.240
Bateke			0.040			0.060	0.040	0.140
Biaka		0.235				0.235	0.294	0.764
Daba			0.050		0.050			0.100
Ewondo		0.020					0.122	0.142
Fali			0.020					0.020
Fulbe Cameroon					0.030	0.030		0.060
Mandara			0.054					0.054
Mbenzele	0.020	0.265				0.061	0.612	0.958
Ngoumba		0.023			0.045		0.068	0.136
Sanga	0.100	0.033	0.167		0.033	0.033	0.033	0.400
Tali			0.050					0.050
Uldeme			0.040					0.040
East Africa								
Mozambique		0.027			0.018			0.045
South-West Africa								
Angola			0.023	0.045	0.091			0.159
Cabinda	0.009	0.018	0.055	0.036	0.118		0.009	0.245
Insular Central Africa								
São Tomé[a]		0.073	0.042	0.010	0.042		0.042	0.209
São Tomé[b]		0.029	0.068		0.078		0.019	0.193
America								
F.B.I. database	0.010	0.003	0.028	0.022	0.036	0.006	0.003	0.108
Choco		0.020	0.061					0.081
Garifuna			0.045					0.045

[a] Unpublished data, J. Rocha.
[b] Trovoada et al. (2004).

L1c1a1 now fall into a clade which we termed L1c5 in order to avoid confusion with the L1c3 haplogroup of Salas et al. (2002). Finally, we identify a new clade, L1c4, which contains mostly sequences from Babinga, a population analyzed in this study for the first time. The TMRCA, the sequence diversity, the mean number of pairwise differences and the mismatch distribution for each clade and for the entire L1c are presented in Fig. 2. It is worthy of note that both L1c and four of its sub-haplogroups (L1c1a, L1c1b, L1c1c and L1c2) show a smooth unimodal distribution (with raggedness values from 0.005 to 0.044), while L1c5 and L1c4 present a multimodal distribution (with raggedness values from 0.054 to 0.089; see Fig. 2).

Our network is more population-structured than that proposed by Salas et al. (2002), presumably because of the inclusion in this dataset of three additional Pygmy populations (Fig. 1). Almost all individuals from the five Western Pygmy populations (184 out of 189; 97% of the total) can be found in the specific sub-haplogroups L1c1a, L1c4 and L1c5, whereas they were only a fraction in distinct clades (L1c*, L1c1a and L1c1a1) of Salas et al. (2002). Pygmies account for 67% of the sequences belonging to L1c1a (43 out of 64), 89% of L1c4 (39 out of 44) and 81% of L1c5 (102 out of 126). This prevalence cannot be simply attributed to a biased composition of the L1c database, since Pygmies represent only 42% of the total sample (189 individuals out of 455). Bantu and African Americans account for most of the variability of the rest of the network; they represent 84% of the sequences belonging to L1c1b and 95% of L1c2. Finally, African-Americans account for 69% of L1c1c sequences.

4. Discussion

4.1. Phylogeography of L1c

Before this study, little was known concerning the mitochondrial heritage of a fundamental anthropological component of Africa, the Pygmies. These are hunter-gatherers living in the equatorial forest which are characterized by a very short stature (~150 cm in males, on average; Cavalli-Sforza, 1986). According to current view

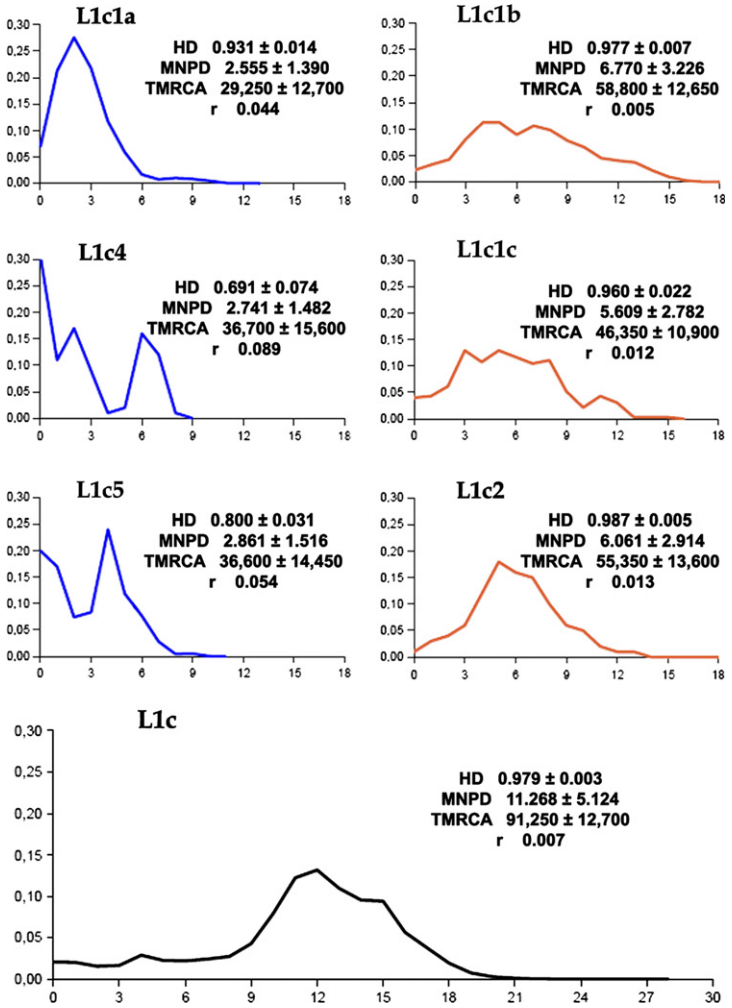

Fig. 2. Mismatch distributions of L1c and its sub-haplogroups. HD, haplotype diversity; MNPD, mean number of pairwise differences; TMRCA, time to the most recent common ancestor; r, raggedness (blue, Pygmy clades; orange, Bantu clades).

most mtDNA variation in Pygmies would reflect admixture with Bantu or a persistence of plesiomorphic characters (Salas et al., 2002). The only previous claim of a Pygmy specific mitochondrial component is represented by an investigation which included two Western Pygmy populations (Mbenzele and Biaka from the CAR). In that study we proposed that a lineage of the L1c haplogroup (L1c1a1 according to the previous nomenclature) is of Pygmy origin (Destro-Bisol et al., 2004a). In the study reported here, it was possible to analyze three additional Western Pygmy populations (Bakola and Baka from Cameroon and Babinga from the Popular Republic of Congo), and to extend the analysis to the hypervariable region 2. This larger dataset and increased resolution level made it possible to substantially revise the L1c phylogeography and gain new insights into the mitochondrial variation of Western Pygmies.

We identified three L1c clades (L1c1a, L1c4 and L1c5, formerly called L1c1a1) which can be considered to be of Western Pygmy origin. The three L1c clades mentioned above contain most of Western Pygmy sequences, from 78% in Baka to 100% in Bakola, while Western Pygmies account for 79% of all sequences assigned to L1c1a, L1c4 and L1c5. These three clades are present in each of the five Western Pygmy populations, with only two exceptions: L1c4 is missing in the sample of Bakola from Cameroon, and L1c5 was not detected among the Babinga from the Popular Republic of Congo. This could be due to the effect of genetic drift. In fact, both these populations have a low haplotype diversity (0.693 ± 0.074 in Babinga and 0.744 ± 0.048 in Bakola).

The quasi-ubiquitous presence of these three L1c clades in the Western Pygmy populations indicates that these groups retain strong signs of their common evolutionary history, despite their small demographic dimensions and the geographical distance separating the Babinga of Congo from the Pygmies from Cameroon and the CAR. On the other hand, it should be noted that no sequence belonging to the L1c clades shared by the Western Pygmy populations studied here, and no other L1c haplotype was detected among the Eastern Pygmies of the Ituri Forest in the Democratic Republic of Congo (Vigilant et al., 1991). Furthermore, the three Western Pygmy groups studied *ex novo* resemble the two populations analyzed by Destro-Bisol et al. (2004a) in that they lack the L1e and L2 haplogroups, which occur at a high frequency among Eastern Pygmies. These results reconfirm a substantial mitochondrial genetic diversity between Western and Eastern Pygmies, as expected from a separation between the two Pygmy groups predating the Bantu expansion (Destro-Bisol et al., 2004a, 2006).

The Pygmy clades L1c1a, L1c4 and L1c5 also contain some sequences of Bantu (Bakaka, Cabinda, Ngoumba, Ewondo, Sanga, Bateke, Bassa) and São Tomean individuals (a total of 40 that account for 17% of the clades mentioned above) and 10 African American individuals from North America and Choco (4%). This finding may be explained by the persistence of a plesiomorphic character and/or gene flow from Pygmies to Bantu. Three lines of evidence favour the second possibility. Firstly, most of the Bantu haplotypes belonging to L1c1a, L1c4 and L1c5 are particularly similar to Pygmy haplotypes: in fact, 9 out of 28 are shared and 12 out of 28 are one-step neighbour. Secondly, most of the Bantu populations mentioned above live or lived in geographical proximity to Western Pygmies. Finally, ethnographic observations and genetic evidence suggest that unions between Pygmy females and Bantu males are favoured by social constraints, whereas taboos make those between Pygmy males and Bantu females difficult (Cavalli-Sforza, 1986; Destro-Bisol et al., 2004b). It is interesting to note that evidence of gene flow of uninlinearly transmitted characters from Western Pygmies to Bantu is limited to mtDNA. In fact, sharing of Y-chromosomal haplogroups between the two population groups is likely due to introgression of Bantu lineages into Western Pygmies (haplogroup E3a) or maintenance of ancestral characteristics diluted elsewhere by more recent demographic events [haplogroups B2a1 and B2b3*x (B2b3a)] (Destro-Bisol et al., 2004b).

Another means to test the hypothesis that L1c1a, L1c4 and L1c5 are of Pygmy origin is to take their sequence variation into consideration. The demographic history of populations is thought to be reflected in various parameters of intrapopulational variation of their mitochondrial pool, including haplotype diversity, mean number of pairwise differences and the mismatch distribution (Harpending et al., 1993). As previously done by Watson et al. (1997), we extended this same line of reasoning to specific mitochondrial lineages. Essentially, we imply that the different demographic histories of Western Pygmies and Bantu should have left a detectable signature in some of their mtDNA clades. L1c4 and L1c5 show a relatively low haplotype diversity and mean number of pairwise differences together with a multimodal mismatch distribution (see Fig. 2). These features are usually observed in populations with a small and constant effective size, as Pygmies are generally considered (Harpending et al., 1993; Von Haeseler et al., 1996; Excoffier and Schneider, 1999). By contrast, L1c1a shows a rather high haplotype diversity (see Fig. 2) and a unimodal mismatch distribution, which can be explained by a recent event of expansion of a Pygmy group of populations that also influenced the history of this clade (Harpending et al., 1993).

The remaining L1c clades—L1c1b, L1c1c and L1c2—show very different features from those discussed above. In fact, they are mostly composed of Bantu and African American individuals (98%), whereas Pygmy presence is limited to two haplotypes (three individuals). L1c1b, L1c1c and L1c2 are characterized by a high haplotype diversity and unimodal mismatch distribution (see Fig. 2), which is typical of populations which have recently expanded (Harpending et al., 1993; Excoffier and Schneider, 1999). All these features suggest that these haplogroups originated in Bantu or pre-Bantu populations.

4.2. Genetic signatures of the prehistory of Central Africa

Current views regarding the peopling of Central Africa suggest that Pygmies should be considered as a group of populations in genetic continuity with the first inhabitants of the tropical forest, while Bantu populations are thought to have spread into this area only 2–3 kya, in the course of their expansion through sub-Saharan Africa (Cavalli-Sforza, 1986). Unfortunately, little is known concerning the relationships between the ancestors of present-day Pygmies and Bantu.

Several archaeological sites have been discovered in the African equatorial belt (Lanfranchi and Clist, 1991; Cornelissen, 2002; Mercader and Martí, 2003). In general terms, analyses of lithic elements of the Middle Stone Age (MSA) in West-Central Africa suggest a cultural evolution along a *continuum* rather than the occurrence of substantial cultural discontinuities in the time period spanning from 100 to 10 kya (Phillipson, 1993; Newman, 1995; Cornelissen, 2002). Paleoclimatological studies based on the analysis of pollen remains from the MSA indicate alternate phases of humid and arid climate between 40 and 12 kya (Sayer et al., 1992; Cornelissen, 2002). These fluctuations probably caused expansion and fragmentation processes of the equatorial forest, a scenario able to generate a fragmented pattern of population distribution and, ultimately, a genetic separation. Therefore, archaeological and paleoclimatological studies seem to have contrasting implications for the peopling of Central Africa, whose relative robustness is difficult to assess.

Another attempt to gain insights into the evolutionary relationships between populations may be based on the phylogenies of unilinearly transmitted polymorphisms, with the caution deriving from the fact that this approach is primarily concerned with the history of genes and not of populations. Prior to this study there was no suitable mitochondrial or Y-chromosomal lineage, which could be used to address the evolutionary issues under consideration. In fact, all haplogroups so far studied have been interpreted as group-specific characters, such as the Y-chromosomal E3a and the mitochondrial L3e2b for Bantu and the Y-chromosomal B and the previous L1c1a1 (now L1c5) for Western Pygmies; otherwise, they have been considered plesiomorphic characters, such as L0a and L2a (Underhill et al., 2001; Salas et al., 2002; Destro-Bisol et al., 2004a; Destro-Bisol et al., 2004b).

The L1c haplogroup of mitochondrial DNA shows the features necessary to attempt a phylogenetically based assessment of the evolutionary relationships between Western Pygmies and Bantu. In fact, it retains a signature of a phase common to the ancestors of the two groups, while encompassing some specific sub-clades which can mark their divergence. This reconstruction can be inserted into a chronological framework using TMRCA estimates, a parameter which gives a measure of the time needed for the creation of the variation observed in a given phylogeny (Forster et al., 1996; Saillard et al., 2000). Even with this approach, demographic aspects of populations must be taken into account in order to avoid wrong or biased inferences. The similarity among L1c TMRCA estimates based on all sequences (91,200 ± 12,700 ya), on Pygmy (99,500 ± 19,700 ya) and Bantu clades (81,300 ± 13,400 ya) indicates that the different demographic histories of the two groups do not substantially influence the results.

Therefore, the L1c TMRCA may be used as a minimum estimate of the phase common to the ancestors of the two groups, while the TMRCAs of the L1c sub-clades may provide an indication of their divergence. Following this principle, our findings allow us to draw a scenario where, at least 90 kya, ancestors of Western Pygmies and Bantu were still forming a genetically coherent hunting-gathering population. Their separation could only have occurred many generations later, between 60 and 30 kya as shown by the TMRCAs of the Pygmy clades. This hypothesis may be supported by current interpretations of archaeological evidence if we assume that genetic separation had not been paralleled by cultural differentiation for some 20–50 kya after the population split. This might have occurred because of a substantial *stasis* of the lithic *repertoires* and/or maintenance of some cultural contacts between the two groups. On the other hand, the cycles of fragmentation and expansion of the tropical forest provide an ecological explanation for the separation between the ancestors of Western Pygmies and Bantu as suggested by the mtDNA data. In fact, the time period of intense ecological change (between 40 and 12 kya) overlaps with that proposed for the genetic separation (between 60 and 30 kya).

Any attempt to shed light on the peopling of Central Africa is made difficult by the incompleteness of the data accumulated so far and the limited value of some more indirect lines of evidence. Nonetheless, the results of our L1c study allowed us to draw up an evolutionary scenario. While it must be considered as a working hypothesis rather than a conclusive account for the reasons discussed above, we achieved the result of proposing a first reference for future studies on the evolutionary relationships between Western Pygmies and Bantu.

Acknowledgments

The research presented here was supported by the European Science Foundation EUROCORES (programme "The Origin of Man, Language and languages") and the Spanish Ministerio de Educación y Ciencia (project "Language, culture and genes in Bantu: a multidisciplinary approach to the bantu-speaking populations of Africa", BFF2002-10206-E). We also acknowledge grants from the M.I.U.R. (Cofin Project 2003054059 "Dna e Biodemografia: approccio integrato allo studio della mobilità umana"), the University of Rome "La Sapienza" (Ateneo project C26A034352 "Analisi della variabilità in geni sottoposti a selezione in popolazioni dell'Africa sub-Sahariana"), and The Portuguese Foundation for Science and Technology (project OCTI/42510/ANT/2001 "The

Anthropogenetics of São Tomé Archipelago: a case study on human microevolution"). We thank Donata Luiselli, Ilaria Boschi, Carlo Colella and Andrea Ferrini for help with mtDNA sequencing, Oscar Lao for help with computing, Giuseppe Carpaneto for useful comments on the manuscript. Last but not least, we sincerely thank all blood donors, whose willingness made this study possible.

Appendix A. Supplementary data

The following material is available online: Figure S1: Control for phantom mutations. Median Joining Network built without considering speedy mutations (Bandelt et al., 2002); Table S1: HVR-1 sequences of the six populations analyzed for the first time in this work (Babinga, Baka, Bakola, Bateke, Ngoumba and Sanga); Table S2: Entire L1c haplotypes dataset used in this work. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2006.09.014.

References

Alves-Silva, J., Santos, M., Guimarães, 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.

Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H., Coulson, A.R., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G., 1981. Sequence and organization 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., Quintana-Murci, L., Salas, A., Macaulay, V., 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am. J. Hum. Genet. 71, 1150–1160.

Beleza, S., Gusmão, L., Amorim, A., Carracedo, A., Salas, A., 2005. The genetic legacy of western Bantu migrations. Hum. Genet. 117, 366–375.

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

Cavalli-Sforza, L.L., 1986. African pygmies: An evaluation of the state of research. In: Cavalli-Sforza, L.L. (Ed.), African Pygmies. Academic Press, Orlando, pp. 361–426.

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara Benerecetti, 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.

Coia, V., Destro-Bisol, G., Verginelli, F., Battaggia, C., Boschi, I., Cruciani, F., Spedini, G., Comas, D., Calafell, F., 2005. Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. Am. J. Phys. Anthropol. 128, 678–681.

Cornelissen, E., 2002. Human responses to changing environments in Central Africa between 40,000 and 12,000 B.P.J.. World Prehist. 16, 197–235.

Destro-Bisol, G., Coia, V., Boschi, I., Verginelli, F., Caglià, A., Pascali, V., Spedini, G., Calafell, F., 2004a. The analysis of variation of mtDNA hypervariable region-1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. Am. Nat. 163, 212–226.

Destro-Bisol, G., Donati, F., Coia, V., Boschi, I., Verginelli, F., Caglià, A., Tofanelli, S., Spedini, G., Capelli, C., 2004b. Variation of female and male lineages in Sub-Saharian populations: the importance of sociocultural factors. Mol. Biol. Evol. 21, 1673–1682.

Destro-Bisol, G., Battaggia, C., Batini, C., Spedini, G., 2006. The Western Pygmies from the Central African Republic: new results on autosomal loci. J. Anthropol. Sci., 84, 161–164.

Excoffier, L., Schneider, S., 1999. Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. Proc. Natl. Acad. Sci. USA 96, 10597–10602.

Gill, P., Jeffreys, A.J., Werrett, D.J., 1985. Forensic application of DNA fingerprints. Nature 318, 577–579.

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

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.

Green, L.D., Derr, J.N., Knight, A., 2000. mtDNA affinities of the peoples of North-Central Mexico. Am. J. Hum. Genet. 66, 989–998.

Harpending, H.C., Sherry, S.T., Rogers, A.R., Stoneking, M., 1993. The genetic structure of ancient human populations. Curr. Anthropol. 34, 483–496.

Ingman, M., Gyllensten, U., 2006. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. Nucl. Acids Res. 34, D749–D751.

Krings, M., Salem, A.H., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., Sajantila, A., Paabo, S., Stoneking, M., 1999. mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? Am. J. Hum. Gen. 64, 1166–1176.

Lanfranchi, R., Clist, B. (Eds.), 1991. Aux origines de l'Afrique Centrale. CCF/CICIBA, Paris.

Mercader, J., Martí, R., 2003. The Middle Stone Age occupations of Atlantic central Africa: new evidence from Equatorial Guinea and Cameroon. In: Mercader, J. (Ed.), Under the Canopy: The Archaeology of Tropical Rain Forests. Rutgers University Press, London, pp. 64–92.

Meyer, S., Weiss, G., von Haeseler, A., 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152, 1103–1110.

Monson, K.L., Miller, K.W.P., Wilson, M.R., Dizinno, J.A., Budowle, B., 2002. The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Sci. Comm. 4. (http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm).

Newman, J.L., 1995. The Peopling of Africa. Yale University Press, New Haven.

Pakendorf, B., Stoneking, M., 2005. Mitochondrial DNA and human evolution. Annu. Rev. Genom. Hum. G. 6, 165–183.

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.

Phillipson, D.W., 1993. African Archaeology. Cambridge University Press, Cambridge.

Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, A., Comas, D., 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum. Genet. 115, 439–447.

Rando, J.C., Pinto, F., Gonzalez, A.M., Hernandez, 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.

Richards, M., Macaulay, V., 2001. The mitochondrial gene tree comes of age. Am. J. Hum. Genet. 68, 1315–1320.

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., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P., Macaulay, V., Carracedo, A., 2002. The making of the African mtDNA landscape. Am. J. Hum. Genet. 71, 1082–1111.

Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., Carracedo, A., 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am. J. Hum. Genet. 74, 454–465.

Salas, A., Richards, M., Lareu, M.V., Sobrino, B., Silva, S., Matamoros, M., Macaulay, V., Carracedo, A., 2005. Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA. Am. J. Phys. Anthropol. 128, 855–860.

Santos, C., Montiel, R., Sierra, B., Bettencourt, C., Fernandez, E., Alvarez, L., Lima, M., Abade, A., Aluja, M.P., 2005. Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: A model using families from the Azores Islands (Portugal). Mol. Biol. Evol. 22, 1490–1505.

Sayer, J.A., Harcourt, C.S., Collins, N., 1992.The Conservation Atlas of Tropical Forests. Africa IUCN, WCMC Macmillan, London.

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

Torroni, A., Miller, J.A., Moore, L.G., Zamudio, S., Zhuang, J., Droma, T., Wallace, D.C., 1994. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. Am. J. Phys. Anthropol. 93, 189–199.

Trovoada, M.J., Pereira, L., Gusmão, L., Abade, A., Amorim, A., Prata, M.J., 2004. Pattern of mtDNA variation in three populations from São Tomé e Principe. Ann. Hum. Genet. 68, 40–54.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon Lahr, M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L., 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. 65, 43–62.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D., Wilson, A.C., 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc. Natl. Acad. Sci. USA 86, 9350–9354.

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.

Von Haeseler, A., Sajantila, A., Paabo, A., 1996. The genetical archaeology of the human genome. Nat. Genet. 14, 135–140.

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