Hum Genet (2004) 115: 439–447
DOI 10.1007/s00439-004-1164-0

ORIGINAL INVESTIGATION

Stéphanie Plaza · Antonio Salas · Francesc Calafell · Francisco Corte-Real · Jaume Bertranpetit · Ángel Carracedo · David Comas

Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola

Received: 27 February 2004 / Accepted: 10 June 2004 / Published online: 1 September 2004

Abstract Africa is the homeland of humankind and it is known to harbour the highest levels of human genetic diversity. However, many continental regions, especially in the sub-Saharan side, still remain largely uncharacterized (i.e. southwest and central Africa). Here, we examine the mitochondrial DNA (mtDNA) variation in a sample from Angola. The two mtDNA hypervariable segments as well as the 9-bp tandem repeat on the COII/tRNA^{lys} intergenic region have allowed us to allocate mtDNAs to common African haplogroups. Angola lies in the southern end of the putative western branch of the Bantu expansion, where it met the local Khoisan populations. Angolan mtDNA lineages show basically a Bantu substrate with no traces of Khoisan lineages. Roughly, more than half of the southwestern mtDNA pool can be assigned to west Africa, ~25% to central Africa and a significant 16% to east Africa, which points to the western gene pool having contributed most to the mtDNA lineages in Angola. We have also detected signals of extensive gene flow from southeast Africa. Our results suggest that eastern and western Bantu expansion routes were not independent from each other, and were connected south of the rainforest and along the southern African savannah. In agreement with historical documentation, the analysis also showed that the Angola mtDNA genetic pool shows affinities with the African lineages from Brazil, the main American destination of the slaves from Angola, although not all lineages in Brazil can be accounted for by the Angolan mtDNA pool.

S. Plaza · F. Calafell · J. Bertranpetit · D. Comas (✉)
Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut I de la Vida, Universitat Pompeu Fabra,
Doctor Aiguader 80,
08003 Barcelona, Spain
e-mail: david.comas@upf.edu
Tel.: +34-93-5422802
Fax: +34-93-5422844

A. Salas · Á. Carracedo
Unidad de Genética, Instituto de Medicina Legal, Universidad de Santiago de Compostela,
Galicia, Spain

F. Corte-Real
Instituto de Medicina Legal, Servicio de Biología Forense,
Coimbra, Portugal

Introduction

Although the pre-colonial history of parts of Africa has been carefully researched, little is known about the southwestern region that forms contemporary Angola as it was before the arrival of the Europeans in the late 1400s. The area currently known as Angola has been inhabited since prehistoric times, and Khoisan people are thought to have been the first settlers of this territory. According to linguistics, one of the greatest expansions that has modelled the African landscape is the dispersal of Bantu languages. Current evidence suggests that the original Bantu homeland was located in the southeastern part of Nigeria (i.e. the Benue valley of southern Nigeria) and/or the northwestern part of Cameroon (i.e. the grasslands of western Cameroon; Newman 1995). The Bantu expansion probably coincided with the end of the Neolithic Age (about 5,000 BP) and was at least, at some stage, related to the diffusion of agriculture and iron metallurgy (Phillipson 1993). The Bantu expansion toward the south split into two major paths: the western route expanding to the south along the Atlantic coast; and the eastern route, passing by over the rainforest, to the area of the Great Lakes, and subsequently to the south. Convergence between eastern and western Bantu routes might have occurred on multiple occasions during different periods of time, although its extent is not well characterized. The settlement of Bantu-speaking farmers in the southwestern side of the subcontinent began around 4,000 BP, when yam-growers with Neolithic tools spread into the rainforest of Cameroon. Farmers speaking western Bantu languages gradually occupied all of central Africa, expanding over different sorts of terrain (coastal routes and through the rainforest). The dispersal was favoured by the adoption of an iron-based technology (presumably accompanying the introduction of new crops). It is believed that the western

Bantu expansion had important consequences on the demography of the native populations, since it marked the first appearance of agriculture, which could have increased the carrying capacity by an order of one magnitude (Ammerman and Cavalli-Sforza 1984). Additionally, it seems that local languages were influenced, and ultimately replaced by Bantu languages (Phillipson 1993; Vansina 1995). Khoisan speakers may have been completely taken over by the Bantu expansion or they may have moved towards the south and centre of the continent, where they still inhabit part of southern Africa, including southern Angola, mostly in harsh environments like the Kalahari desert.

In colonial times, the Portuguese started slave trading in the African Atlantic coast buying slaves from African chiefs to work in sugar plantations in São Tomé and subsequently sent them to America. Until the Portuguese abolished the slave trade, Angola was the source of as many as two million slaves for the Americas (Thomas 1997). More than half of these went to Brazil, a third to the Caribbean, and from 10 to 15% to the Río de la Plata area on the southeastern coast of south America. As a result of the slave trade, the Angolan territory may have lost around 4 million people (Thomas 1997).

The African origin of modern humans is widely accepted, a fact that increases the interest of genetic knowledge on African populations, which not only show a higher heterogeneity than any other geographical region, but also a complex population history that genetics is helping to unravel. A large compilation of classical marker data across the African continent (Cavalli-Sforza et al. 1994) showed a clear differentiation between north African and sub-Saharan populations [first principal component (PC)], the relationship between Ethiopian and Khoisan populations (second PC), and the similarity between Bantu populations (third and fourth PC). However, greater resolution has been achieved with Y-chromosome and mtDNA data, which are both uniparental

Table 1 HVSI and HVSII sequences found in Angola (*HPG* haplogroup, + presence of the 9-bp deletion, *ND* not determined)

	1111111111111111111111111111111111111111111111111111111111111111	0000111111111122222222233333	HPG	9bp
	6666666666666666666666666666666666666666666666666666666666666666	6799445558888990003446990011		del
	0000011111111111111111122222222222222222222223333333333333333333	4335360122569580476473679956		
	3568912223446677888888901122233456666777889990001112245556669	..		
	8186344691586823235789293535604940345018673451491690724590280	12		
ANDERSON	AATTTCTTGTGCACTCAACCCTCTGACCAACTACTCACTCCCACCCTATAGCCTCCTCTTG	CAAAGTCCTCGCATCATGTAGACA---G		
AN3	....C...A..T.TC....TGC....T..G.........T..G.....C..T.C.....	..GC....C.A.G.....C.AG..--C.	L0a1a	-
AN94	....C...A..T.TC....TGC....T..G.........T..G.....C..T.....	..GC......A.G.....C.AG..--C.	L0a1	-
AN26	........A..T.TC....TGC....T..G.................C..T.....	T.G.......A.G..G..C.AG..--C.	L0a1	nd
AN37	..........T..C....TAC....T..G.................C..T.....	T.G.....C...G...CAC.AG..CCC.	L0a2	+
AN47	..........T..C....TGC....T..GT................C..T.....	T.GC....C...G.....C.AGT.C-C.	L0a2	+
AN68	..........T..C....TGC....T..G.................C..T.....	T.G.....C...G...CAC.AG..--C.	L0a2	-
AN75	.......C..........T.C....T........T.T.T........C.....	.G.....TCTT..C......AG..--C.	L1b	-
AN105	.......C..........T.C....T........T.T.T........C.....	.G......CTT..C......AG..--C.	L1b	-
AN5	G.................T.C....T...........T..GT....C........T...	.G.....TCT.ACCT.....AG.G--CA	L1c1	-
AN25	....C...A.........T.C....T.......C....T..GT....C........T.C.	.G.....TCT.ACC......AG.G--CA	L1c1	-
AN29	........A.........T.C....T.........C..TG..T....C.......CT...	.G.....TCT.ACC......AG.G--CA	L1c2	-
AN27	........A.........T.C....T.........C..TG..T....C.A......T...	.G.....TCT.ACC......AG.GC-CA	L1c2	nd
AN64	........A.........T.C....T.........C..TG..T....C.......CT...	.G.....TCT.ACC......AG.G--CA	L1c2	-
AN88	........A.A.......T.C..A.T.........C.CTG..T....C........T...	.G.....TCT.ACCT.....AG.G--CA	L1c2	nd
AN46	....C...A........C...C...GT...........T...T....C...T..T.T...	.G.....TCT.AC.......AG..C-CA	L1c3	nd
AN92	........A...G......T.C....T.....G......T.........C.....	.G...C..CT...CT.....AG..--C.	L1e	-
AN12	..........................T...........T.T.T..............A	.G..AC..CT...C.....G..C-C.	L2a	nd
AN125	..........................T...TC......T...TT..............A	.G..AC.......C.....G..C-C.	L2a	nd
AN69	...C.....................T...........T...T...G...........A	.G..AC..C....CT......G..--C.	L2a1	-
AN72	.........C.......C...C....TTG.T........T...T...G...........A	.G...C..C....C.....G..C-C.	L2a1	-
AN17	.........C...........C....TT..T.......T.T...T...G...........A	.G...C..C...........G..CCC.	L2a1	nd
AN86	....................CT...T...........T...T...G...........A	.G...C..C....C.....G..C-C.	L2a1	-
AN45	....................CT...T...........T...T...G...........A	.G...C..C....C.....G..C-C.	L2a1	-
AN71	....................CT...T...........T...T...G...........A	.G...C..C....C.....G..C-C.	L2a1	-
AN23	..........................T...........TT..T...G...........A	.G...C..C....C.....G..--C.	L2a1a	nd
AN57	..C.......................T...........TT..T...G...........A	.G...C..C....C.....G..--C.	L2a1a	-
AN40	..........................T...........TT..T.T.G...........A	.G...C..C....C.....G..C-C.	L2a1a	-
AN73	.....A..A..............A.T...........T..............T.....A	.G...CT.CT...CT.....G..--C.	L2b	-
AN100	.....A..A..............A.T...........T..............T.....A	.G...CT.CT...CT.C....G..C-C.	L2b	nd
AN1	....C.C..........C...C....T...........T...................C..	.G..................G..--C.	L3b	nd
AN28	......C...................T.....................A..........	.G....T.C...........G..--C.	L3d1	nd
AN52	......C.........CC...C....T...........T......C.C..........	.G......C...........G..--C.	L3d3	-
AN7	........A........C...C....T......T.................T........	.G....T.....G..G.....G..CCC.	L3e1	+
AN60	........A........C...C....T......T.................T........	.G....T.....G..G.....G..--C.	L3e1	+
AN9	..........................T.....................T........	.G....T.....G..G.....G..C-C.	L3e1	-
AN54	..................T.......T.....................T........	.G....T.....G..G.....G..C-C.	L3e1a	-
AN74	..................T.......T.....................T........	.G....T.....G..G.....G..C-C.	L3e1a	-
AN130	..............CTCC...C....T.....................T........	nd	L3e2b	nd
AN2	..........................T.........T...........C.....	.G....T......C.......G..--C.	L3e3	-
AN90	....C.....................T.........T....................	.G....T......C.......G..--C.	L3e3	-
AN111	....C.....................T.........T....................	.G....T......C.......G..--C.	L3e3	-
AN42	......................C..T.....................C.....	nd	L3f	-
AN121	.....T...................T...............T.....CG.....T..C..	.G...C..C....C.....G.G..--C.	L3g	-
AN53	.G...T..............CT...T...............T.....CG.....T..C..	.G...C..C....C.....G.G..--C.	L3g	-

markers with highly resolved phylogeographies. Since Vigilant et al. (1991), numerous African populations have been surveyed for mtDNA variation, whereas fewer studies have focused on the global African Y-chromosome variation (Scozzari et al. 1999; Underhill et al. 2000, 2001). The Y-chromosome landscape in sub-Saharan Africa has been characterized by haplogroups A, B, and part of haplogroup E (E3a) (nomenclature from the Y Chromosome Consortium 2002), the last one related to the Bantu expansion and dated around 3,000–5,000 years ago (Thomas et al. 2000).

The African mtDNA landscape is dominated by lineages belonging to L haplogroups [L0, L1, L2 and L3A (Chen et al. 1995; Watson et al. 1997; Salas et al. 2002)]; other African specific non-L haplogroups are M1 [with an east African origin (Quintana-Murci et al. 1999)], and U6 [specific from north Africa (Rando et al. 1998; Plaza et al. 2003)]. Several L-mtDNA lineages are present in the Khoisan mtDNA pool [L0d and L0k (Bandelt and Foster 1997)], while others seem to have been dispersed along sub-Saharan Africa by Bantu farmers (Salas et al. 2002): L0a (Bandelt et al. 1995; Chen et al. 1995), L3b (Watson et al. 1997), L2, L3e, and L1e (Alves-Silva et al. 2000; Bandelt et al. 2001). The COII/tRNA^{lys} intergenic 9-bp deletion related to part of the L0a haplogroup has also been suggested as an important Bantu marker (Soodyall et al. 1996). Salas et al. (2002) confirmed these findings by an analysis of the African mtDNA variation as a whole, but missing out the Angolan region. On the base of the composition of the Brazilian lineages available, Salas et al. (2002) speculated that the western Bantu expansion might have involved more assimilation of indigenous lineages in the forest zone (mainly in the form of L1c lineages) than the eastern stream. They also postulated the existence of four major founders of both west and east African origin involved in the eastern Bantu expansion (L0a1a, L0a2, L2a1a and L2a1b).

In the African study of Salas et al. (2002), the overall mtDNA composition for the continent was described, except the uncharacterized southwest. The analysis of Angolan mtDNA could shed light on four main issues related to the African genetic diversity. First, the characterization of the mtDNA gene pool of southwestern Africa, of which little is known. Second, establishing the extent of the Bantu demographic expansion in its western part and the possible admixture with Khoisan lineages. Third, determining the degree of differentiation between both Bantu expansion routes (west and east) by comparison with the southeast region. Fourth, determining the contribution of southwest Africa to the mtDNA pool of Brazil, the main American destination of Angolan slaves.

Material and methods

Samples, mtDNA amplification and sequencing

A total of 44 unrelated individuals were sampled: 43 Mbundu, the second largest population in Angola (making up one-quarter of the total population), and one Bakongo (AN9). All the individuals had maternal ancestors from Angola. They were analysed for both hypervariable segments I (HVSI) and II (HVSII) of the mtDNA control region, and for the COII/tRNA^{lys} 9-bp intergenic deletion (positions, 8281–8289; Anderson et al. 1981) (Table 1). Total DNA was extracted using a Chelex method (Lareu et al. 1994).

Both hypervariable segments were amplified in one reaction using primers L15996 and H408 (Vigilant et al. 1989), and the amplified product was purified with the Gene Clean kit (BIO 101). Both strands were sequenced using the primers L15996 and H16401 for HVSI, and L29 and H408 for HVSII (Vigilant et al. 1989). Both hypervariable segments were sequenced with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), and the sequence products were run on an ABI PRISM 377 sequencer (Applied Biosystems).

The 9-bp tandem repeat (CCCCCTCTA) of the COII/tRNA^{lys} intergenic region was amplified by PCR using the primers and methods described by Comas et al. (2004). The amplified product was run on an automatic sequencer ABI PRISM 377 and the fragment sizes were analysed with the GeneScan software analysis package.

Phylogenetic and population analyses

Sequences from positions 16024 to 16391 and 63 to 322 (according to Anderson et al. 1981) were used in the present analysis and are available on the following web site (http://www.upf.edu/cexs/recerca/bioevo). HVSII sequences were not determined for two individuals, due to a lack of DNA. The information provided by both the HVSI and HVSII was used to classify the sequences into haplogroups according to Salas et al. (2002) (cf. Chen et al. 1995; Watson et al. 1997; Rando et al. 1998; Quintana-Murci et al. 1999; Alves-Silva et al. 2000; Bandelt et al. 2001; Pereira et al. 2001; Torroni et al. 2001). Following the suggestion by Richards and Macaulay (2000) and Mishmar et al. (2003), L1 nomenclature has been changed according to the sequence scheme that appeared in Salas et al. (2004a).

Haplotype genetic diversity (G) was calculated as $G=[n/(n-1)](1-\Sigma_{I=1}^{k} p_I^2)$, where p is the frequency of each of the k different sequences in the sample, with the Arlequin 2.000 program (Schneider et al. 2000). In order to compare the present results with those from other populations, data for the first mtDNA hypervariable region (positions 16090–16365) from a number of sub-Saharan population samples were taken from the literature (Table 2). Several samples from Mozambique (Pereira et al. 2001; Salas et al. 2002) were pooled. Sequences from Cabo Verde islands (Brehm et al., 2002) were also considered as a single population. Other sub-Saharan and American samples described by Salas et al. (2002) were also used in parts of the analysis, such as sequence sharing. For some analyses, population samples were grouped into major geographic areas: west, east, central,

southeast and southwest Africa (Table 2, Fig. 1). L sequences from Brazil (Alves-Silva et al. 2000) were used as an additional population sample throughout the analyses.

Population genetic structure was tested through analysis of molecular variance (AMOVA) (Excoffier et al. 1992), using the Arlequin 2000 program (Schneider et al. 2000). A spatial analysis of the molecular variance (SAMOVA) was also performed using the SAMOVA 1.0 program (Dupanloup et al. 2002) by presetting different numbers of population groups. This approach defines groups of populations that are geographically homogeneous and maximizes the proportion of total genetic variance due to differences between groups.

A correspondence analysis was performed from haplogroup absolute frequencies using the SPSS package. Correspondence analysis provides a method for representing frequency data in an Euclidian space so that the results can be visually examined for structure (Greenacre 1992). For data in a typical two-way contingency table, both the row variables (populations) and the column variables (haplogroups) are represented in the same space, allowing the gathering of the relationships not only within row or column variables but also between row and column variables.

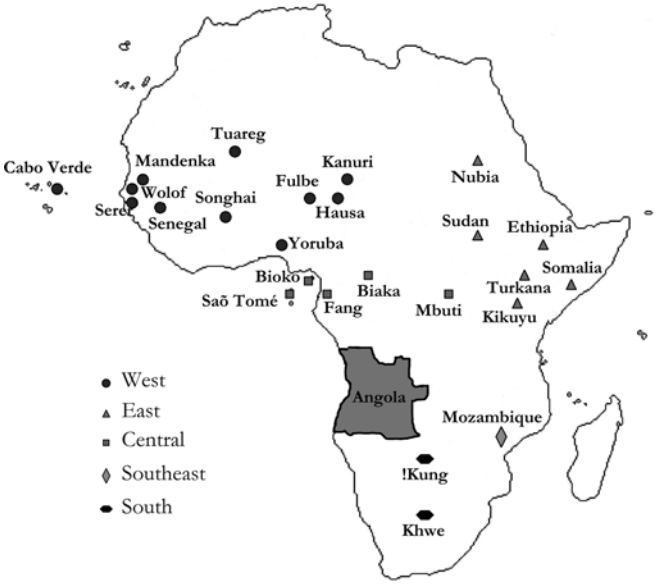

Fig. 1 Location of the samples used for reference and their regional ascription

Table 2 Samples used in the present study

Populations	Geographic region	Code	Sample size	References
Angola	Southwest	An	44	Present study
Mozambique	Southeast	Mz	416	Pereira et al. (2001); Salas et al. (2002)
Cabo Verde	West	CV	292	Brehm et al. (2002)
Mandenka	West	Mn	119	Graven et al. (1995)
Fulbe	West	Fu	61	Watson et al. (1996)
Hausa	West	Ha	20	Watson et al. (1996)
Kanuri	West	Ka	14	Watson et al. (1996)
Songhai	West	So	10	Watson et al. (1996)
Tuareg	West	Tg	26	Watson et al. (1996)
Yoruba	West	Yo	35	Watson et al. (1996); Vigilant et al. (1991)
Senegalese	West	Sn	50	Rando et al. (1998)
Serer	West	Sr	23	Rando et al. (1998)
Wolof	West	Wo	48	Rando et al. (1998)
Sudan	East	Su	76	Krings et al. (1999)
Nubia	East	Nu	80	Krings et al. (1999)
Kikuyu	East	Ki	25	Watson et al. (1996)
Somali	East	Sm	27	Watson et al. (1996)
Turkana	East	Tk	37	Watson et al. (1996)
Ethiopian	East	Et	74	Thomas et al. (2000)
Bubi	Central	Bu	45	Mateu et al. (1997)
São Tomé	Central	ST	50	Mateu et al. (1997)
Fang	Central	Fa	11	Pinto et al. (1996)
Mbuti	Central	Mb	20	Vigilant et al. (1991)
Biaka	Central	Bk	17	Vigilant et al. (1991)
!Kung	South	Kg	67	Vigilant et al. (1991); Chen et al. (2000)
Khwe	South	Kw	31	Chen et al. (2000)
Brazil	America	Br	69	Alves-Silva et al. (2000)

Results

Angolan mtDNA genetic composition

The genetic diversity found in Angola for HVSI (0.992 ±0.007), HVSII (0.982±0.009) and both hypervariable regions together (0.997±0.006) is similar to the diversity found in other sub-Saharan samples analysed previously (Salas et al. 2002, 2004a). All the sequences obtained in the present analysis can be assigned to the African specific L lineages, such as L0 (13.6%), L1 (22.7%), L2 (29.5%), and L3A (34.1%), and have been classified into haplogroups according to Salas et al. (2002, 2004a).

Within L0, which includes at least four haplogroups (L0a, L0d, L0k and L0f), only sequences belonging to haplogroup L0a (with the L0a1 and L0a2 subclades), have been found in our sample from Angola. The Khoisan L0d and L0k, and the eastern African lineage L0f are not found in the present sample. The presence in Angola of the eastern African L0a1 subclade, which constitutes one tenth of the lineages found in east Africa, might be due to migration from east/southeast Africa. On the other hand, a central origin was proposed for the L0a2 subclade and has been associated with the COII/tRNA^{lys} 9-bp deletion (Soodyall et al. 1996). Two of the three L0a2 sequences found in Angolans carried the intergenic COII/tRNA^{lys} 9-bp deletion. Therefore, the presence of L0a2 sequences in the southwest (i.e. Angola) and southeast [i.e. Mozambique (Pereira et al. 2001; Salas et al. 2002)] might be explained by migration from central Africa (see also discussion in Salas et al. 2002).

Haplogroup L1b is known to be most frequent in west Africa (13%), and is also present in African Americans (10%) as a consequence of the African slave trade. Diffusion of this haplogroup seems to have been very limited in southwest Africa since it is represented only by two individuals (4.5%) in Angola (both matching western but not eastern African sequences).

The presence of the ancient haplogroup L1e is mainly limited to east Africa at a low frequency (4%). It is rare in other parts of Africa: only the subclade L1e2 has been found in two Mozambicans, one Mbuti, and one Egyptian. The single L1e sequence type found in Angola has no match with the rest of L1e African sequences.

The L1c haplogroup was postulated to have originated in central Africa towards the Atlantic Coast (Salas et al. 2002), since it has been observed at relatively high frequencies in central Africa as well as in African Americans [up to almost 23% in Brazilians (Alves-Silva et al. 2000)], but it is rare elsewhere. As predicted, L1c lineages are frequent in Angola (16%), with three sublineages being represented; L1c1 (4.5%), L1c2 (9.1%), and L1c3 (2.3%). Nonetheless, the network of the L1c sequences of sub-Saharan, African American (Fig. 5a in Salas et al. 2002), and Angolan lineages (data not shown), locates Angolan L1c lineages at the tips of the branches. This seems to suggest that southwest Africa is not the homeland of L1c. No matches were found between Angolan and other African L1c sequence types.

Two of the four subclades of haplogroup L2 were present in Angola: L2a, the most common and widespread L2 subclade in Africa, accounting for 25% of the Angolan lineages, and L2b (6.8%). L2a1a (and L2a1b) have undergone dramatic expansion in southeast Africa (Salas et al. 2002), a demographic event not detected in our Angolan L2a1a sequences. Many Angolan L2a mtDNAs match southeast African types, although most of the time they also match western sequences. L2b Angolan types match three Mozambique sequences (Pereira et al. 2001; Salas et al. 2002) as well as two !Kung individuals (Chen et al. 2000) southern types. Some of these southeast/southwest sequence matches could indicate recent gene flow between both regions, rather than a provenance from a common source in western or eastern Africa.

The lineage L3A includes at least L3b, L3d, L3e, L3f, and L3g haplogroups (cf. Salas et al. 2002). L3e is the second most frequent haplogroup in Angola with a frequency ~21%. Four subclades of L3e are found in Angola: L3e1 (6.8%), L3e1a (4.5%), L3e2b (2.3%), L3e3 (6.8%). L3e is not very abundant in east Africa (~3%), but it is more prevalent in west (~11%) and central (~20%) Africa. Eastern Bantu expansion could have carried L3e at significant frequencies to southeast (~15%), and south (~11%) Africa; whereas the western Bantu route could have carried L3e western (L3e2b and L3e3) or west–central (L3e1) types to Angola.

The haplogroups L3g/L3f appear to have an East African origin (Salas et al. 2002). Thus, these lineages could have reached Angola directly from their original source but also carried by the western Bantu expansion, since L3g/L3f have been detected along the African Atlantic coast from Cameroon to Angola (Salas et al. 2004b).

Sequence sharing

Angolan sequences were compared with the sequence dataset included in Salas et al. (2002) and the sample from Cabo-Verde reported by Brehm et al. (2002). Most of the Angolan mtDNAs were already found in other sub-Saharan populations. We examined how many identical HVSI sequences (positions 16090–16365) were shared between southwest Africa and the rest of the African regions, as well as with American sequences. Variation at positions 16182–16185 and length polymorphism at the polyC were not considered. A total of 20 Angolan HVSI sequences were found in other African regions, being southeast Africa (data pooled from Pereira et al. 2001 and Salas et al. 2002) the region with the highest number of matches with southwest Africa. From these haplotypes found, most of them were also found in other sub-Saharan populations, and just a few are only present in southwest, southeast and America: one sequence (AN125, belonging to L2a) was shared only by southwest and southeast; three (AN94, AN9, and AN54/74; belonging to L0a1, L3e1, and L3e1a respectively) were found in these two regions and in America, and three haplotypes (AN37, AN5, and AN53; belonging to L0a2, L1c1, and L3g respectively) were shared only between southwest and America. Only one sequence (AN40, belonging to L2a1a) was present only in west and southwest regions.

Angola within the African mtDNA landscape

In order to place Angolans within the sub-Saharan African mtDNA framework, a correspondence analysis based on the absolute haplogroup frequencies was performed using the African populations analysed (Table 2). The Khoisan (!Kung and Khwe) and the Pygmy (Biaka and Mbuti) samples distort the correspondence analysis (data not shown) due to the different haplogroup composition (high frequencies of L0d and L0k) and sequence ambiguities of Khoisan and Pygmy samples, respectively. These were, therefore, excluded from subsequent analyses. Only sequences belonging to the major sub-Saharan haplogroup L and M1 lineages were included in this analysis.

The correspondence analysis is shown in Fig. 2. The first dimension (47.2%) separates the southeastern populations from the rest of Africa, a pattern also observed by Salas et al. (2002). These samples appear isolated at one edge of the plot characterised by the L0a1a, L0a2, L0d, L1c2, L2a1a, L2a1b, L3e1, and L3e3 lineages. The second dimension (38.7%) shows a clear separation between western and eastern populations. The eastern populations are mainly associated with lineages such as L0a1, L1e, L3a, L3g, and M1. Central and southwestern populations are placed between the populations from other geographi-

cal areas. Southwest Africans are situated between eastern and western African samples, but closer to the latter. This suggests a contribution from both Bantu expansion routes to the Angolan genetic pool. A principal component analysis based on the relative haplogroup frequencies and a multidimensional scaling (MDS) analysis based on *Fst* distances between sequences were also performed. These also displayed similar results (data not shown).

Genetic variation and population structure

An analysis of molecular variance (AMOVA, Table 3) was performed on the sub-Saharan populations used for the correspondence analysis. When all populations were considered as a single group, 5.60% ($P<0.001$) of the genetic variance was found between populations, showing significant genetic heterogeneity among these populations. In order to ascertain how this genetic structure was partitioned, different grouping criteria were applied. When linguistic affiliation was considered (Niger-Kordofanian/Afro-Asiatic/Nilo-Saharan/Portuguese-Creole), a non-significant 0.27% of the variance was attributed to differences among linguistic groups. This means, that the genetic diversity is not structured according to a linguistic classification. When the populations were roughly classified into west (western, central and southwestern samples in Table 2) and east Africa (eastern and southeastern samples in Table 2), 4.17% ($P<0.05$) of the genetic variance was attributable to geographic groups, whereas 3.12% ($P<0.001$) was due to differences among populations from the same geographical area. When the geographical area was defined more precisely (western, eastern, central, southwestern, and southeastern Africa), the variance attributable to geographic groups increased to 4.84% ($P<0.001$), and the differences between populations within the same geographical area decreased to 1.73% ($P<0.001$). Finally, if Angolans were grouped with the central Africans, according to their position in the correspondence analysis, the differences among geographical areas increased to 4.93% ($P<0.001$), and differences among populations within groups decreased to 1.68% ($P<0.001$). This suggests a close genetic relationship between Angolans and central Africans.

A SAMOVA was performed with the sub-Saharan African samples used in the AMOVA both in order to define groups of populations that are geographically adjacent and genetically homogeneous, and to maximise the proportion of genetic variance between them (Table 3). When two groups were sought, the maximum proportion of total genetic variance between groups (5.13%, $P<0.001$) was found between the whole set of western populations (Table 2) plus São Tomeans, and the rest of populations. When the number of groups is set to four or five, Angola and Mozambique constitute one such group. The absence of any intervening sampled population makes Angola and Mozambique topological neighbours and allows the SAMOVA algorithm grouping them, which highlights their genetic relationship. This result stresses the affinity between the southeast and southwest regions, both at the end of the Bantu expansion routes.

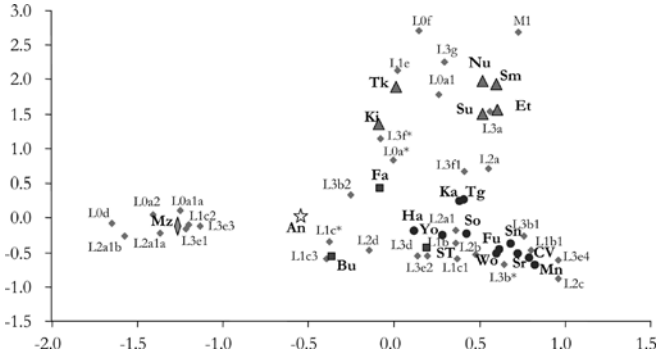

Fig. 2 Two-dimension plot of the analysis of the correspondence based on the absolute L and M1 haplogroup frequencies of the west, east, central, southwest and southeast African populations. *Symbols* as in Fig. 1 and *abbreviations* as in Table 2

Discussion

Angola, and the southwestern part of Africa in general, has been up until now a missing piece in the African genetic puzzle. Here we have studied an Angolan sample to address four main issues: (1) the characterisation of the southwest African gene pool, (2) the estimation of admixture between Bantu and Khoisan populations, (3) the location of Angola in the mtDNA African landscape, and (4) the exploration of the putative geographical origin of the African lineages detected in Brazil, the main Angolan slave trade destination.

Characterisation of the southwestern Africa gene pool

Correspondence and phylogeographic analyses show that Angolan mtDNAs mirror the west and central African gene pools, with a minor eastern African component. Moreover, AMOVA has shown that the amount of genetic variation between groups is higher when Angolans are considered together with central Africans.

One of the most intriguing aspects of the Angolan gene pool is the high frequency of L1c haplogroup (15.9%). These lineages are also frequent in Brazilians and other Afro-American samples [more than one-third of L1c haplotypes reported belong to African Americans (Salas et al. 2002)]. They have also been found in high proportions (~22%) in central Africa. L1c is much rarer elsewhere: from 0% in south Africa to ~5% in southeast Africa. A putative Angolan (Alves-Silva et al. 2000) or central African (Salas et al. 2002) origin for the haplogroup L1c has been postulated. Angolan L1c sequences lie at the tips of the L1c phylogeny (Fig. 5a, Salas et al. 2002) far from the root sequence. However, none of the L1c Angolan sequence types matches those described in African samples. In fact, matches for L1c sequences in African Americans have been mainly found in central Africa. The

Table 3 Analysis of the molecular variance (AMOVA) and spatial analysis of the molecular variance (SAMOVA) in sub-Saharan populations (abbreviations as in Table 2)

	Among groups	Among populations within groups	Within populations
AMOVA			
All populations		5.60**	94.40**
Linguistic affiliation	0.27 NS	5.42**	94.31**
Geographical area (West vs. East)	4.17*	3.12**	92.71**
Five geographical areas (W, SW, E, SE and Central)	4.84**	1.73**	93.43**
Four geographical areas (W, E, SE and SW+Central)	4.93**	1.68**	93.39**
SAMOVA			
1. Western populations+ST	5.13**	2.52**	92.35**
2. The rest of populations			
1. Western populations+ST	5.22**	2.23**	92.55**
2. Sm+Et			
3. Ki+Su+Nu+Tk+An+Mz+Fg+Bu			
1. Western populations+ST	5.40**	1.56**	93.04**
2. Eastern populations+Fg			
3. Bu			
4. An+Mz			
1. Western populations+ST	5.63**	1.30**	93.07**
2. Bu			
3. Su+Sm+Et			
4. Ki+Nu+Tk+Fg			
5. An+Mz			

*P<0.05
**P<0.001

central African populations sampled so far are the islands of Bioko and São Tomé, both inhabited, respectively, by an old western Bantu isolated group and descendants of slaves (Mateu et al. 1997). In the mainland, populations studied so far are a small Fang sample (n=10) from Equatorial Guinea, and two Pygmy samples from the central African Republic and the Congo Democratic Republic. Thus, most of mainland central Africa remains to be sampled, including countries as large as Cameroon, Gabon and the Republic of the Congo. A very recent report (Destro-Bisol et al. 2004) shows that most of the mtDNA sequences in a western Pygmy population from Cameroon, the Mbenzele, belong to L1c (96.4%, being 9.1% L1c*, 29.1% L1c1a*, and 58.2% L1c1a1). The authors suggest a local origin for the L1c1a1 offshoot. In summary, the heartland of L1c may lie in the still largely uncharacterized coastal facade of central Africa (from Cameroon to the Republic of Congo), which may correspond to a secondary focus of the Bantu expansion, from where it may have been in part exported to the Americas. These results confirm the role of L1c in the southwestern Bantu expansion suggested by Salas et al. (2002) based on the Brazilian L-mtDNA composition.

Lack of Khoisan component in southwest Africa

Khoisan people might have occupied a vast territory before the Bantu expansion, which gradually displaced or assimilated Khoisan speakers. As predicted by Alves-Silva et al. (2000), none of the Khoisan characteristic lineages (L0d or L0k) were found in the southwest mtDNA pool. In addition, note that L0d and L0k have not been found in the large African-American survey by Salas et al. (2004a). In the extant Khoisan groups, !Kung and Khwe (Chen et al. 2000), L0d and L0k haplotypes constitute around 36% and 24% of the lineages respectively. The probability of not finding a particular lineage that is present in a population at a frequency of f in a sample of size n is given by $\alpha=(1-f)^n$. Therefore, the maximum contribution of Khoisan lineages to Angolans is compatible with the observation that the absence of L0d and L0k in a sample of 44 Angolans would be less than 10.8% (with a confidence of P=0.05), which is evidence for a dramatic (and almost complete) replacement of the Khoisan maternal lineages by the Bantu people. However, given the different carrying capacities associated with the hunter-gatherer (Khoisan) and farmer (Bantu) lifestyles, it is expected that, even if the Bantu absorbed all the local Khoisan people, the latter would not have contributed much to the admixture. A larger sample is needed in order to obtain sufficient power to discriminate between the two extreme hypotheses (no Khoisan admixture vs complete assimilation). If the African lineages found in Brazilians are taken as proxies of the Angolan mtDNA pool (having non-significantly different haplogroup frequencies, Fisher's exact test, P=0.164), they could be pooled to increase the sample size to 113. In that case, the maximum possible contribution of Khoisans to the extant Angolan mtDNA pool would drop to 4.3%. Therefore, the present gene pool of Angolans is basically the result of the Bantu expansion within the region with a very small or non-existent contribution by the Khoisan sequences. Additionally, no L0d sequences were detected in the African-American sequences analysed by Salas et al. (2004a). This also indicates the lack of L0d in southwest Africa, since

Angolan slaves were also carried to other American destinations.

Most of the lineages that are thought to have been dispersed by the Bantu are found in SW Africa, such as L0a1, L0a2, L3b, L3e, L2a1a, and L3e (Bandelt et al. 1995; Chen et al. 1995; Watson et al. 1997; Pereira et al. 2001; Salas et al. 2002). This fact, added to the lack of Khoisan lineages in Angola, points to a basically Bantu substrate of the extant Angolan gene pool. Thus, the Bantu expansion was clearly more demic (in the sense of population replacement) in the southwest than in the southeast, where remnants of ancient settlers (related to extant Khoisan) are obvious.

Angola in the African mtDNA landscape

The southwest African lineages seem to have originated mostly from west/central Africa. The analysis of the molecular variance shows a clear grouping between southwest and the central Africa region. This result suggests a large contribution of the western stream of the Bantu expansion after dispersion and assimilation of indigenous lineages in the equatorial zone. In addition, the correspondence analysis displayed a clear separation between west and east Africa, but place the southwest region in an intermediate position between west, east and southeast Africa. This suggests that the western and eastern Bantu expansions were not independent events, but that they most likely joined below the tropical forest zone and then dispersed through the southern areas of Africa. This is also supported by the spatial analysis of the molecular variance (SAMOVA, Table 3), where Angola and Mozambique are jointly clustered in all the analyses and constitute a separate cluster when samples are divided in four or more groups. Besides their geographic proximity, this reveals a genetic homogeneity between both regions and suggests that they shared a common set of haplogroups brought about by the Bantu expansion. A local differentiation of the lineages followed by gene flow between both regions may have also helped to maintain a close relationship between both areas. These results are also in accordance with the African geographic landscape, since the equatorial forest seems to have acted as a strong genetic barrier and limited the interaction between both western and eastern Bantu streams. On the other hand, the southwest and southeast areas are separated by the savannahs, easier to cross and more densely populated than the tropical rainforest, therefore, more permeable to gene flow.

Alternatively, the Bantu people may have reached both southwestern and southeastern Africa from a common origin elsewhere. This could explain some common shared lineages, but it is still hard to reconcile with the historical and linguistic data.

Following the same phylogeographic approach employed in Salas et al. (2002), west Africa would have contributed to ~60% of the southwestern mtDNA composition, central Africa to 23% and east Africa to a significant 16%.

African mtDNA contribution to the Brazilian population

As Angola was known to provide the major number of African slaves to Brazil (Thomas et al. 1997), inferences on the Angolan mtDNA composition were done on the basis of the Brazilian mtDNA pool (Alves-Silva et al. 2000). Our results confirm the prediction of Alves-Silva et al. (2000). Brazilian and Angolan samples share a low number of sequences, as they both displayed high genetic diversity (0.992±0.007 for Angolans and 0.994±0.004 for Brazilians). When the Brazilian population was introduced as a single group in the correspondence analysis based on haplogroup frequencies (data not shown), Brazil and central Africa are clustered as a single group. Therefore, although haplogroup frequencies in Brazil are roughly similar to those in Angola, the African mtDNA pool of the Brazilian population did not come exclusively from Angola, but is most likely the result of admixture of African slaves from different colonies distributed in west, southwest, and central Africa (Guinea Coast, Saõ Tomé; see also Salas et al. 2004a).

Acknowledgements We would like to thank Mònica Vallés, Òscar Lao, and Gemma Berniell, Universitat Pompeu Fabra, for technical assistance, suggestions, and careful reading of the manuscript. We would also like to thank three anonymous reviewers for their constructive comments. The present study was supported by the Dirección General de Investigación, Ministerio de Ciencia y Tecnología, Spain (BOS2001-0794 and BFF2002-10206-E), Direcció General de Recerca, Generalitat de Catalunya (2001SGR00285), and Ministerio de Sanidad y Consumo (FIS 2003/PF012). The research presented in this article is supported by the framework of the European Science Foundation EUROCORES programme "The Origin of Man, Language and Languages". S.P. received a fellowship from the Direcció General de Recerca, Generalitat de Catalunya (2000FI00696). A.S. is supported by the Isidro Parga Pondal program.

References

Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461

Ammerman AJ, Cavalli-Sforza LL (1984) The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

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

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Bandelt HJ, Alves-Silva J, Guimarães PE, Santos MS, Brehm A, Pereira L, Coppa A, Larruga JM, Rengo C, Scozzari R, Torroni A, Prata MJ, Amorim A, Prado VF, Pena SD (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet 65:549–563

Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. Ann Hum Genet 66:49–60

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) History and geography of human genes. Princeton University Press, Princeton

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (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 YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (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

Comas D, Plaza S, Wells RS, Yuldasheva N, Lao O, Calafell F, Bertranpetit J (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 12:495–504

Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglià A, Pascali V, Spedini G, Calafell F (2004) 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

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11:2571–2581

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

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

Greenacre M (1992) Correspondence analysis in medical research. Stat Methods Med Res 1:97–117

Krings M, Salem AE, Bauer K, Geisert H, Malek AK, 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 MV, Phillips CP, Carracedo A, Lincoln PJ, Syndercombe Court D, Thomson JA (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

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

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100:171–176

Newman JL (1995) The peopling of Africa. Yale University Press, New Haven

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, 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 DW (1993) African archaeology. Cambridge University Press, Cambridge

Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60:321–330

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D (2003) Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. Ann Hum Genet 67:312–328

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

Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt HJ (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 (2000) Genetic data and the colonization of Europe: genealogies and founders. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research, Cambridge, pp 139–151

Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, Sánchez-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 MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A (2004a) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465

Salas A, Torroni A, Richards M, Quintana-Murci L, Hill C, Macaulay V, Carracedo A (2004b) The phylogeography of mtDNA haplogroup L3 g in Africa and the Atlantic slave trade. Am J Hum Genet (in press)

Schneider S, Kueffer JM, Roessli D, Excoffier L (2000) Arlequin (ver 2.000): a software environment for the analysis of population genetics and Biometry Lab. University of Geneva, Geneva

Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellito D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martínez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A (1999) Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet 65:829–846

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608

Thomas H (1997) The slave trade-the history of the Atlantic slave trade:1440–1870. Macmillan, London

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000) Y chromosomes 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, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69:1348–1356

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62

Vansina J (1995) New Linguistic evidence and the "Bantu expansion" J Afr Hist 36:173–195

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (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 AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S (1996) MtDNA sequence diversity in Africa. Am J Hum Genet 59:437–444

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

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348