AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 140:302–311 (2009)

mtDNA Variability in Two Bantu-Speaking Populations (Shona and Hutu) From Eastern Africa: Implications for Peopling and Migration Patterns in Sub-Saharan Africa

Loredana Castrì,[1]* Sergio Tofanelli,[2] Paolo Garagnani,[1] Carla Bini,[3] Xenia Fosella,[2] Susi Pelotti,[3] Giorgio Paoli,[2] Davide Pettener,[1] and Donata Luiselli[1]

[1]*Dipartimento di Biologia Evoluzionistica Sperimentale, Unità di Antropologia, Università di Bologna, I-40126 Bologna, Italy*
[2]*Dipartimento di Biologia, Unità di Antropologia, Università di Pisa, Pisa, Italy*
[3]*Dipartimento di Medicina e Salute pubblica, Sezione di Medicina Legale, Università di Bologna, I-40126 Bologna, Italy*

KEY WORDS genetic variability; mtDNA haplogroups; sub-Saharan Africa; Bantu speaker migrations

ABSTRACT In this study, we report novel data on mitochondrial DNA in two of the largest eastern Bantu-speaking populations, the Shona from Zimbabwe and the Hutu from Rwanda. The goal is to evaluate the genetic relationships of these two ethnic groups with other Bantu-speaking populations. Moreover, by comparing our data with those from other Niger-Congo speaking populations, we aim to clarify some aspects of evolutionary and demographic processes accompanying the spread of Bantu languages in sub-Saharan Africa and to test if patterns of genetic variation fit with models of population expansion based on linguistic and archeological data. The results indicate that the Shona and Hutu are closely related to the other Bantu-speaking populations. However, there are some differences in haplogroup composition between the two populations, mainly due to different genetic contributions from neighboring populations. This result is confirmed by estimates of migration rates which show high levels of gene flow not only between pairs of Bantu-speaking populations, but also between Bantu and non-Bantu speakers. The observed pattern of genetic variability (high genetic homogeneity and high levels of gene flow) supports a linguistic model suggesting a gradual spread of Bantu-speakers, with strong interactions between the different lines of Bantu-speaker descent, and is also in agreement with recent archeological findings. In conclusion, our data emphasize the role that population admixture has played at different times and to varying degrees in the dispersal of Bantu languages. Am J Phys Anthropol 140:302–311, 2009.

A widely investigated topic in African genetic studies is the expansion of agriculturalist people speaking languages belonging to the Bantu branch of the Niger-Congo linguistic family (Pereira et al., 2001; Salas et al., 2002; Plaza et al., 2004; Beleza et al., 2005), a crucial demographic event in the history of the continent (Ehret, 1981; Vansina, 1995; Holden, 2002; Phillipson, 2003). According to the most widely accepted model deriving from linguistic and archeological data, desertification pressure caused by climatic changes around 4,000 years ago caused some Bantu-speaking farmers to move from their putative homeland in western-central Africa (somewhere near the confluence of the Niger and Benue rivers) across central Africa along eastern and western directions that brought them via complex routes to the southernmost part of Africa (Newman, 1995).

Archeological evidence suggests that the two streams of expansion moved southward in different times and manners: the first along the western coast, the second towards the eastern interlacustrine area, then southward along the coast and through the central internal regions (Phillipson, 1993). It is not yet clear if the two waves of advance intermingled along the way, but some contacts with each other (mainly in the central rainforest region) and with local populations have been supposed (Huffman, 1982; Newman, 1995).

The picture coming from linguistic data is complex and is still a matter of debate. Briefly, two main models have been proposed: a wave model implying that "there have been many successive Bantu-speaker dispersals rather than a single continuous expansion" (Vansina, 1995); a continuous or tree model supporting a single "great" expansion of farmers from western-central Africa (Holden, 2002). An intermediate model, with modified trees showing periods of interaction between different

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: University of Bologna.

*Correspondence to: Loredana Castrì, Dipartimento di Biologia Evoluzionistica Sperimentale, Unità di Antropologia, Università di Bologna, via Selmi 3, 40126 Bologna, Italy.
E-mail: loredana.castri@unibo.it

Received 9 January 2008; accepted 17 February 2009

DOI 10.1002/ajpa.21070
Published online 7 May 2009 in Wiley InterScience (www.interscience.wiley.com).

lines of descent, has been postulated by Ehret (2001). The hypotheses on the spread of Bantu languages also differ in the definition of the core region from which the expansion started. Some authors have proposed an earlier radiation from western-central Africa (Cameroon) with subsequent branching in the central rainforest areas (Congo-Kinshasa) and successive westward and east/southeastward radiations (Vansina, 1995; Ehret, 2001). According to other authors the main Bantu-speaker radiation started from western-central Africa after an early split of western and eastern sub-groups (Holden, 2002).

Several studies have provided evidence for the diffusion of genes associated with the Bantu-speaker migrations. It seems that the distribution of many mtDNA lineages can be associated with the spread of Bantu-speaking farmers from western Africa through the central region to the south. Most of these markers apparently spread out along the eastern route of Bantu-speaker migrations (Soodyall et al., 1996; Bandelt et al., 2001; Pereira et al., 2001; Salas et al., 2002), while only a few lineages are believed to have been carried along the western route (Salas et al., 2002; Plaza et al., 2004; Beleza et al., 2005), essentially because of the scarce genetic information on southwestern African populations. On the other hand, studies on Y-chromosome variability have proposed an ancient Bantu-speaker haplotype that originated somewhere near the putative Bantu-speaker homeland and was dispersed by farming populations through the sub-Saharan region (Thomas et al., 2000; Pereira et al., 2002).

In the present study, we analyze genetic variability in two of the largest Bantu-speaking groups, the Shona from Zimbabwe and the Hutu from Rwanda. The mtDNA variability in these groups has not been studied thus far, except in a small sample of Shona from Mozambique (n = 18, Salas et al., 2002). By comparing our new data with a dataset of 1,500 individuals from 37 Niger-Congo-speaking populations, we aim (i) to further contribute to understanding of demographic and evolutionary processes accompanying the spread of Bantu languages in sub-Saharan Africa and (ii) to test if the observed pattern of genetic variation fits with the hypothesis of Bantu-speaker expansion based on archeological and linguistic data.

MATERIALS AND METHODS

Subjects

The Shona people (6.225 million, World FactBook 2006) speak a language related to the southeastern Bantu branch (Ethnologue Classification: Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Southern, Narrow Bantu, Central, S, Shona). They probably moved to the present-day Zimbabwe during the Bantu-speaker expansions between 500 and 1000 A.D. (Phillipson, 1976). The first inhabitants of Zimbabwe, nomadic Khoisan groups, were enslaved and partly absorbed by the Bantu-speaking populations. Today, the Shona people inhabit Zimbabwe, southern Zambia, and western-central Mozambique. All the Shona tribes are similar in language and livelihood (based on agriculture and animal husbandry) but differ in religious beliefs and customs. Each tribe includes different clans characterized by a deep sense of unity and self-identity.

Oral swabs were collected from 59 unrelated individuals speaking a Shona language (Zezuru dialect), all workers at the public Luisa Guidotti Hospital in Mashonaland East, a northeastern region of the Mutoko district (Zimbabwe) (see Fig. 1). DNA was extracted using the GENTRA kit extraction method, following the manufacturer's protocols (Gentra Systems, Minneapolis).

Fig. 1. Geographical location of Shona and Hutu samples.

The Hutu are the largest Bantu-speaking group from Rwanda and Burundi (~85% of the total population, about 15 million people, World FactBook 2006). They speak a language related to eastern Bantu languages (Ethnologue Classification: Niger-Congo, Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Southern, Narrow Bantu, Central, J, Rwanda-Rundi). It has been suggested that they arrived from central Africa through the Congo region, settled around Lake Kivu between the 5th and 11th century A.D. and gradually took over the autochthonous population, the Pigmoid Twa. Between the 14th and 15th century, a group of Hamitic herders, the Tutsi, reached the Kivu region from eastern Africa (Liesegang et al., 1979). At the time of contact, the two ethnic groups were presumably differentiated by physical appearance, cultural traits, and subsistence economy (the Tutsi depended on cow herding, the Hutu on agriculture). However, they soon joined in a single state and shared the same language (*kinyarwanda*) and a common set of religious beliefs. The 42 Hutu samples were collected among the workers of the Catholic parish at Nyarurema in northeastern Rwanda. DNA was isolated from eyebrows as previously described (Tofanelli et al., 2003).

The ethnic origin of all individuals was ascertained by oral interview. The collection of biological samples and biodemographic information was performed in accordance with the ethical standards of the Universities of Bologna and Pisa and with the American Association of Physical Anthropologists Code of Ethics. All donors gave their informed oral consent prior to being included in the research and personal data were treated anonymously.

MtDNA analysis

DNA samples from the 101 individuals were used for mtDNA amplification and sequencing. PCR amplification

of the first hypervariable segment (HVSI) was performed using the primers L15996 and H16401 (Vigilant et al., 1991). To reduce ambiguities in sequence determination, the forward and reverse primers were used to sequence both strands of HVSI using the BigDye Terminator Cycle Sequencing Kit (ver. 1.1, PE Applied Biosystem) according to the manufacturer's protocol. PCR products were purified by Centrisep (PE Applied Biosystem) and sequencing products were separated by capillary electrophoresis with the ABI Prism 310 Genetic Analyzer (Perkin Elmer). Sequences from position 16024 to 16383 were aligned and compared with the reference sequence (rCRS; Andrews et al., 1999) using the Sequence Navigator computer program (Applied Biosystem, Sequence Navigator version 1.0.1). To ensure data quality, all sequences were aligned and edited by two researchers independently. The final consensus sequence was then generated by comparing the two independent results. An additional amplification was performed in case of ambiguities and/or poor quality sequences (as in the case of C-stretch length heteroplasmy).

RFLP typing of coding sites diagnostic for haplogroup assignments was performed by restriction endonuclease analysis of PCR-amplified mtDNA fragments and the restriction fragments were resolved through electrophoresis in 3:1 Nu-Sieve agarose gel. The RFLP analysis was carried out by a hierarchical approach: all individuals were screened for +10806 HinfI, +16389 HinfI, and −3592 HpaI which define respectively the three major African mtDNA haplogroups L0/L1, L2, L3 (Bandelt et al., 2001; Torroni et al., 2001), and then samples were tested for diagnostic RFLP sites which define sub-haplogroups within each major cluster. Namely, individuals lacking the 3592 HpaI site were tested for 10084 TaqI (L3b), 8616 MboI (L3d), and 2349 MboI (L3e); individuals harboring +10806 HinfI were typed for 4310 AluI (L0a), 7055 AluI (L1b/c), and for the presence of the COII/$tRNA^{lys}$ 9-bp deletion (L0a2); individuals with the 16389 HinfI site gain were typed for 13803 HaeIII (L2a), 4157 AluI (L2b), 3693 MboI (L2d), and 13957 HaeIII (L2c) (Bandelt et al., 2001; Torroni et al., 2001). All individuals not assigned to L1, L2, or L3 sub-haplogroups were screened for +10397 AluI (Hg M) and +10871 MnlI (Hg N) sites (Macaulay et al., 1999). Finally, each mtDNA was assigned to the corresponding haplogroup on the basis of the combined RFLP status and HVSI mutational motif, according to the most recent nomenclature system (Salas et al., 2002; Kivisild et al., 2004; Torroni et al., 2006; Behar et al., 2008). In particular, individuals showing the combined HVSI/RFLPs motif 16223, 16293T, 16311, 16355, 16362/−3592HpaI were assigned to Hg L4g (previously reported as L3g, Bortolini et al., 2004; Salas et al., 2004) and individuals showing the motif 16129, 16166, 16187, 16189, 16223, 16278/+10806HinfI were assigned to HgL5 (Kivisild et al., 2004).

Statistical and phylogenetic analyses

Standard diversity indices were calculated by means of Arlequin 3.01 software (Excoffier et al., 2005). Comparisons were made with 37 published Niger-Congo-speaking populations (Supporting Information Table S1) to place the Shona and Hutu samples within the context of the variation observed in their linguistic group. HVSI control region sequences considered for the statistical and phylogenetic analyses range between 16090 and 16370 np. Standard diversity indices, pairwise F_{ST} genetic distances (under the Kimura 2p model, α = 0.26), the statistics necessary for the analysis of molecular variance (AMOVA, Excoffier et al., 1992), were computed with Arlequin, version 3.01 (Excoffier et al., 2005). STATISTICA 6.0 (Statsoft Inc.) was used for Multidimensional scaling (MDS) analysis of pairwise F_{ST} distances and for Correspondence Analysis (CA) of haplogroup frequencies. A spatial genetic analysis was performed using the SGS software (Degen et al., 2001). Genetic distograms, representing mean genetic distances between all pairs of individuals belonging to a spatial distance class plotted against the spatial distance classes, were calculated using the genetic distance of Nei (1972). The reference value indicate distance values expected for a spatially random distribution of haplotypes; distances below the reference were obtained when individuals geographically close are genetically more similar than expected (positive spatial structure), while distances above the reference were obtained when proximal individuals are more divergent than expected (negative spatial structure).

Gene flow rates between pairs of populations were estimated by means of the software Migrate (Beerli and Felsenstein, 2001) which uses a coalescent theory approach for giving maximum likelihood estimates for the number of migrants per generation, N_em (where N_e is the effective population size and m is the migration rate), of n populations. Migration rates were estimated as averages across three independent runs, which included 10 short chain (10,000 genealogies per chain) and three long chains (100,000 genealogies per chain) with increments of 20 and 200 steps. Gene flow rates were assumed to be symmetrical between pairs of populations.

RESULTS

MtDNA analysis and haplotype sharing

We obtained 101 mtDNA HVSI sequences (Table 1), embracing a total of 74 polymorphic sites and 64 different haplotypes. Haplotype diversities were 0.978 (± 0.008) and 0.979 (± 0.010) for the Shona and Hutu samples respectively. In the Shona, the highest frequencies are observed for L0a1a and L0a2 (Table 1 and Fig. 2), both present at fairly high frequencies in other Bantu-speaking populations (Supporting Information Table S3). Sub-lineages L3e2b, quite rare in the southeast region, reach frequencies comparable with those of western Bantu-speakers; indeed, all but one of the L3e2 haplotypes match western sequences (Supporting Information Table S4). At the same time, haplogroups frequently found in southeastern Bantu-speakers, like L3e3, L3e1 and L2a1b, were not detected or occur at low frequencies in the Shona from Zimbabwe. Haplogroups L0k and L0d, typical of Khoisan populations (Salas et al., 2002), are also present among the Shona, although at low frequencies (1.69% each); their presence could reflect gene flow from the pre-existing southern populations, as reported also for other south-eastern Bantu-speaking populations (Salas et al., 2002). An important finding is the high frequency in the Hutu, but not in the Shona, of L0f, a haplogroup common in East Africa (Supporting Information Table S3) (Salas et al., 2002; Kivisild et al., 2004; Tishkoff et al., 2007; Castrì et al., 2008). Also the presence of one L3x1, three L4g, and two L5a2 sequences, lineages mainly restricted to northeastern and eastern

TABLE 1. HVSI sequences and haplogroup frequencies in the Shona and Hutu sample. HVSI mutated positions are reported minus 16,000

Sample	HVSI	RFLPs	Hg	Shona, *n* (%)	Hutu, *n* (%)
RW01	093 148 168 172 187 188G 189 223 230 287 293 311 320	+10806 HinfI −4310 AluI	L0a1a		3 (7.14)
SH01/ RW04	129 148 168 172 187 188G 189 223 230 278 293 311 320	+10806 HinfI −4310 AluI	L0a1a	4 (6.78)	2 (4.76)
RW06	148 172 173 187 188G 189 223 230 311 320 399	+10806 HinfI −4310 AluI	L0a2		1 (2.38)
RW07	111A 148 172 188A 189 223 230 311 320	+10806 HinfI −4310 AluI 9bp del	L0a2		1 (2.38)
SH05	148 172 187 188G 189 223 230 311 320	+10806 HinfI −4310 AluI 9bp del	L0a2	4 (6.78)	
SH09	148 172 187 188G 189 223 230 234 311 320	+10806 HinfI −4310 AluI 9bp del	L0a2	1 (1.69)	
SH10	093 148 172 187 188G 189 223 230 311 320	+10806 HinfI −4310 AluI 9bp del	L0a2	4 (6.78)	
SH14	093 148 172 174 187 188G 189 214 223 230 289 311 320	+10806 HinfI −4310 AluI 9bp del	L0a2	1 (1.69)	
SH15	093 129 148 172 187 188G 189 223 230 311 320	+10806 HinfI −4310 AluI 9bp del	L0a2	1 (1.69)	
SH16	126 148 172 187 188 G 189 223 230 311 320	+10806 HinfI −4310 AluI 9bp del	L0a2	1 (1.69)	
SH17	129 150 166del 172 187 189 212 223 243 265 311	+10806 HinfI −4310 AluI	L0d	1 (1.69)	
SH18	147 166C 172 187 189 214 230 278 291A 311	+10806 HinfI −4310 AluI	L0k1	1 (1.69)	
RW09	129 169 172 186 187 189 223 230 278 311 327 368	+10806 HinfI −4310 AluI	L0f		2 (4.76)
RW11	129 169 172 173 187 189 223 230 239 278 311 327 368	+10806 HinfI −4310 AluI	L0f		4 (9.52)
RW15	129 169 172 173 187 189 210 223 230 239 278 311 327 368	+10806 HinfI −4310 AluI	L0f		1 (2.38)
SH19	126 187 189 223 264 270 278 311	+10806 HinfI −7055 AluI	L1b	2 (3.39)	
SH23	129 187 189 214 223 265 T 278 286A 291 294 311 360	+10806 HinfI −7055 AluI	L1c2	1 (1.69)	
SH24	129 187 189 223 265 C 278 286G 294 311 359 360	+10806 HinfI −7055 AluI	L1c2	1 (1.69)	
SH25	071 129 145 187 189 213 223 234 265C 278 286G 294 311 360	+10806 HinfI −7055 AluI	L1c2	1 (1.69)	
RW08	093 129 145 187 189 213 223 265C 278 286G 294 311 360	+10806 HinfI −7055 AluI	L1c2		1 (2.38)
SH21	129 153 183C 189 215 223 278 294 311	+10806 HinfI −7055 AluI	L1c3	1 (1.69)	
SH22	093 129 183C 189 215 223 278 294 311 360	+10806 HinfI −7055 AluI	L1c3	1 (1.69)	
RW36	111 129 148 166 187 189 223 231 233 239 254 278	+10806 HinfI	L5a2		1 (2.38)
RW37	111 129 148 166 187 188 189 223 254 278 311 360	+10806 HinfI	L5a2		1 (2.38)
SH26	223 234 249 278 292 294 295 390	+16389 HinfI +13803 HaeIII	L2a	1 (1.69)	
SH27	129 223 278 294 390	+16389 HinfI +13803 HaeIII	L2a	1 (1.69)	
RW18	183C 189 223 229 278 291 294 311 368 390	+16389 HinfI +13803 HaeIII	L2a		1 (2.38)
SH28	182C 183C 189 192 223 278 290 294 309 390	+16389 HinfI +13803 HaeIII	L2a1b	1 (1.69)	
SH29	183C 189 223 278 290 294 309 390	+16389 HinfI +13803 HaeIII	L2a1b	1 (1.69)	
SH30/ RW17	189 223 278 294 309 390	+16389 HinfI +13803 HaeIII	L2a1	1 (1.69)	1 (2.38)
SH31	223 234 278 294 309 363A 390	+16389 HinfI +13803 HaeIII	L2a1	1 (1.69)	
SH37	051 223 264 291 294 309 390	+16389 HinfI +13803 HaeIII	L2a1	1 (1.69)	
RW16	223 224 278 294 309 390	+16389 HinfI +13803 HaeIII	L2a1		1 (2.38)
SH32	223 278 286 294 309 390	+16389 HinfI +13803 HaeIII	L2a1a	2 (3.39)	
SH34	183C 189 223 278 286 294 309 390	+16389 HinfI +13803 HaeIII	L2a1a	1 (1.69)	
SH35	129 223 278 286 294 309 390	+16389 HinfI +13803 HaeIII	L2a1a	1 (1.69)	
SH36	111 223 278 286 294 309 390	+16389 HinfI +13803 HaeIII	L2a1a	1 (1.69)	
SH38	114 129 209 213 223 278 354 390	+16389 HinfI +4157 AluI	L2b	1 (1.69)	
RW19	114A 129 213 223 278 355 362 390	+16389 HinfI +4157 AluI	L2b		1 (2.38)
SH39	223 224 278 311 390	+16389 HinfI -13957 HaeIII	L2c	1 (1.69)	
RW20	129 189 278 300 311 354 390	+16389 HinfI -3693 MboI	L2d1		1 (2.38)
RW26	124 223 278 362	−3592 HpaI +10084 TaqI	L3b		2 (4.76)
RW21	093 124 188 223 278 362	−3592 HpaI +10084 TaqI	L3b		1 (2.38)
RW22	093 124 223 278 362	−3592 HpaI +10084 TaqI	L3b		3 (7.14)
SH40/ RW25	124 223 278 311 362	−3592 HpaI +10084 TaqI	L3b2	1 (1.69)	1 (2.38)
SH41	124 223 278 293 311 362	−3592 HpaI +10084 TaqI	L3b2	2 (3.39)	
SH43	124 182C 183C 189 223 278 304 311	−3592 HpaI +10084 TaqI	L3b2	1 (1.69)	
SH44/ RW28	124 223 319	−3592 HpaI +8616 MboI	L3d1	4 (6.78)	1 (2.38)
SH48	223 327	−3592 HpaI +2349 MboI	L3e1	3 (5.08)	
SH51	185 209 223 327	−3592 HpaI +2349 MboI	L3e1a	1 (1.69)	
RW29	185 223 311 327	−3592 HpaI +2349 MboI	L3e1a		1 (2.38)

TABLE 1. (Continued)

Sample	HVSI	RFLPs	Hg	Shona, *n* (%)	Hutu, *n* (%)
SH52	223 311 320	−3592 HpaI +2349 MboI	L3e2	2 (3.39)	
SH54	111 124 223 311 320	−3592 HpaI +2349 MboI	L3e2	1 (1.69)	
SH55	172 189 223 320	−3592 HpaI +2349 MboI	L3e2b	3 (5.08)	
SH58/RW30	172 183 C 189 223 320	−3592 HpaI +2349 MboI	L3e2b	1 (1.69)	1 (2.38)
RW31	051 223 264	−3592 HpaI +2349 MboI	L3e4		1 (2.38)
RW33	129 223 293T 311 343 355 362 399	−3592 HpaI	L4g		1 (2.38)
RW34	093 223 293T 311 355 362 399	−3592 HpaI	L4g		1 (2.38)
RW35	223 293T 311 362 399	−3592 HpaI	L4g		1 (2.38)
SH59	093G 223 287A 293T 301 311 355 362	−3592 HpaI	L4g	1 (1.69)	
RW32	169 207 223 278	−3592 HpaI	L3x1		1 (2.38)
RW38	093 129 220 223 254 311 316 362	−3592 HpaI	L3*		1 (2.38)
RW39	223 311 354 399	−3592 HpaI	L3*		3 (7.14)
RW42	223 316	−3592 HpaI	L3*		1 (2.38)
Total				59	42

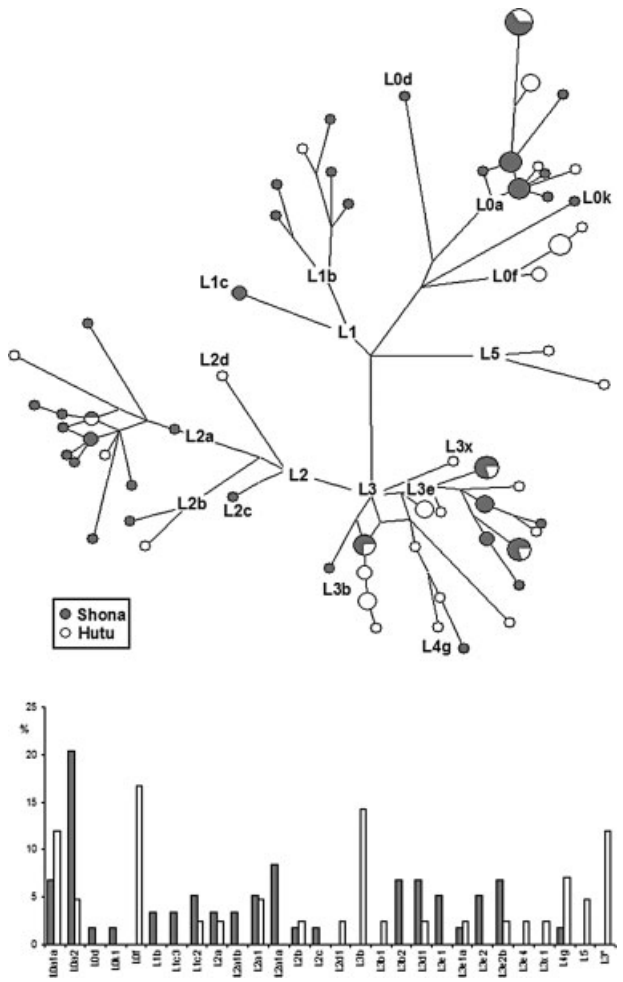

Fig. 2. Haplogroup frequency distributions and Median Joining Network of the Shona and Hutu mtDNAs.

Africa (Supporting Information Table S3), could be due to recent gene flow with populations from these regions.

The different distribution pattern in the Shona and in the Hutu is well represented in the CA (see Fig. 3) which is based on haplogroup frequencies and allows the visualization of the relationships not only within populations but also between populations and alleles (i.e. mtDNA haplogroups). The scatter-plot shows a clear separation of the eastern Bantu-speakers Kikuyu, Sukuma and Hutu from the other populations due to haplogroups L4g, L5, L0f, and L3x. Haplogroups L1c1, L1c2, and L0a1 are associated with the Shona sample, clustering near western Bantu-speaking populations along with two other Mozambican samples (Chwabo and Nyungwe).

Comparison with other African populations

Figure 4 displays a MDS plot based on the matrix of F_{ST} genetic distances between 39 Niger-Congo speaking populations. No clear separation between different Bantu-speaking populations is detectable. Nevertheless, the eastern Bantu-speaking Hutu and Sukuma are placed in a cluster distinct from the southeastern groups which are located in the lower right side of the graph, and western-central and southwestern populations are located on the left side of the graph, with the exception of the Bakaka and Bamileke samples which cluster with SE populations and the Bubi and Herero group which are located on the lower and upper part of the graph, respectively. This distribution pattern could indicate a west-east gradient. To further investigate this hypothesis we performed a spatial genetic analysis using the SGS software (Degen et al., 2001). The resulting distograms (Fig. 5a) showed that there is no evidence of a gradient if we consider all the 39 Niger-Congo speaking population examined in the MDS analysis. When we take into account only 29 Bantu-speaking populations (Fig. 5b), we observed a first increase of genetic distances between populations as spatial distances increase, followed by a negative peak in correspondence of the 1,800 km distance class and a successive further increase. This pattern could be due to evolutionary forces, like gene flow or drift, acting on the genetic cline expected on the basis of the isolation-by-distance model (IBD) (Bertorelle and Barbujani, 1995).

To further investigate the nature and extent of differences among Niger-Congo populations, we computed some parameters of genetic diversity on a database of about 1,500 published HVSI sequences (Supporting Information Tables S1 and S2). A reduction of haplotype diversity can be observed for southwestern and southeastern Bantu-speakers (Table 2 and Supporting Information Table S2), partly reflected also in the mean number of pairwise differences (MNPD). Pairwise-difference distributions of all populations are clearly

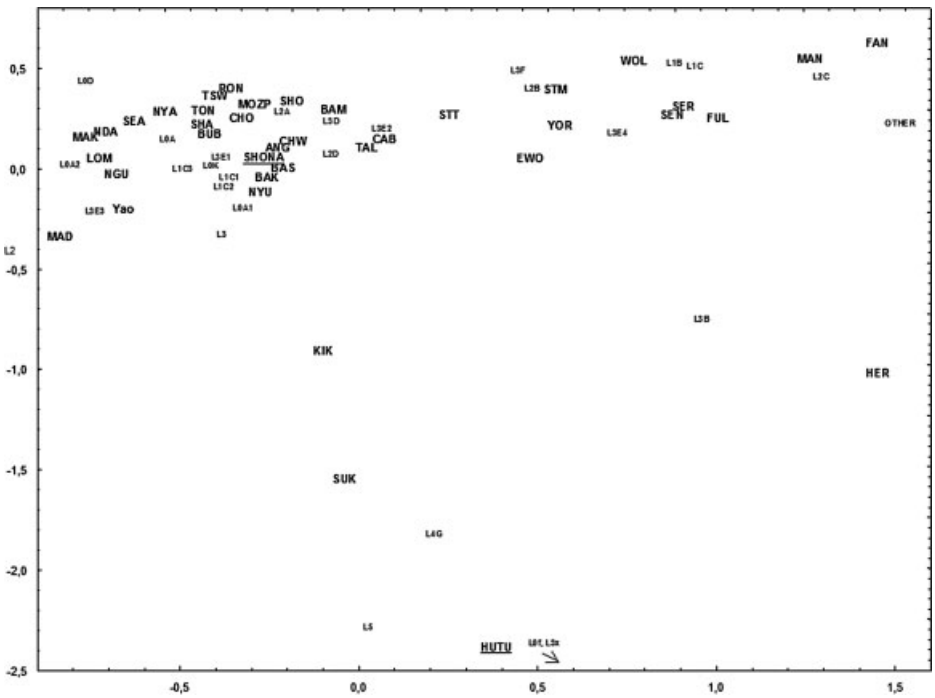

Fig. 3. Correspondence analysis of haplogroup frequencies in 39 Niger-Congo speaking populations. Labels indicate the position of populations (in bold) and haplogroups. Population codes are as follows: FUL = Fulbe; MAN = Mandenka; SEN = Senegal; SER = Serer; WOL = Woloff; YOR = Yoruba; BAM = Bamileke; BUB = Bubi; EWO = Ewondo; FAN = Fang; BAK = Bakaka; BAS = Bassa; STM = São TomémM; STT = São TométT; TAL = Tali; KIK = Kikuyu; *HUTU* = Hutu; SUK = Sukuma; ANG = Mbundu; CAB = Cabinda; HER = Herero; CHO = Chopi; CHW = Chwabo; LOM = Lomwe; MAK = Makhuwa; MAD = Makonde; MOZ = Mozambicans; NDA = Ndau; NGU = Ngoni; NYA = Nyanja; NYU = Nyungwe; RON = Ronga; SEA = Sena; SHA = Shangaan; *SHONA* = Shona; SHO = Shona 2; TON = Tonga; TSW = Tswa; YAO = Yao.

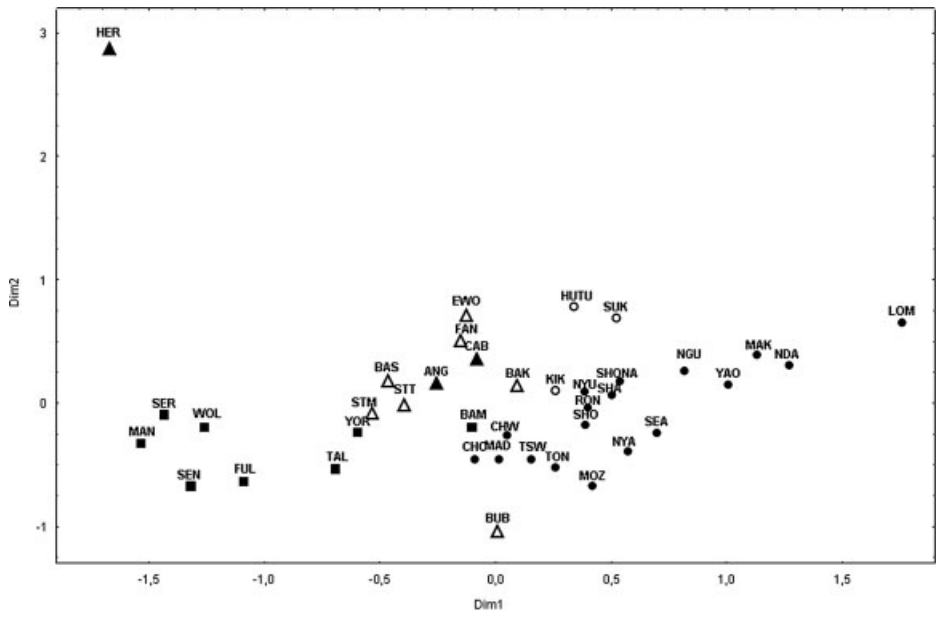

Fig. 4. Multidimensional scaling of F_{ST} distances between 39 Niger Congo-speaking populations. Stress value for MDS = 0.12527 (under the 1% one-tail cut-off value, Sturrock and Rocha, 2000). Populations labels as in Fig. 3. Black squares = western Niger-Congo populations; white triangles = west-central Bantu; black triangles = south-western Bantu; white circles = eastern Bantu; black circles = south-eastern Bantu.

bell-shaped. Only the Hutu sample shows a slight bimodal distribution, but the sum of square deviations (SSD) is not significant and the Harpending's raggedness index was small and not significant. Tajima' D and Fu' *F*s statistics give low negative values for almost all populations. The *P*-values indicated that the D statistics are not significantly lower than the values that would be expected under equilibrium. The *F*s statistics are all significant, indicating a population expansion (Excoffier and Schneider, 1999).

The apportionment of genetic variation between and within populations was assessed by AMOVA; the popula-

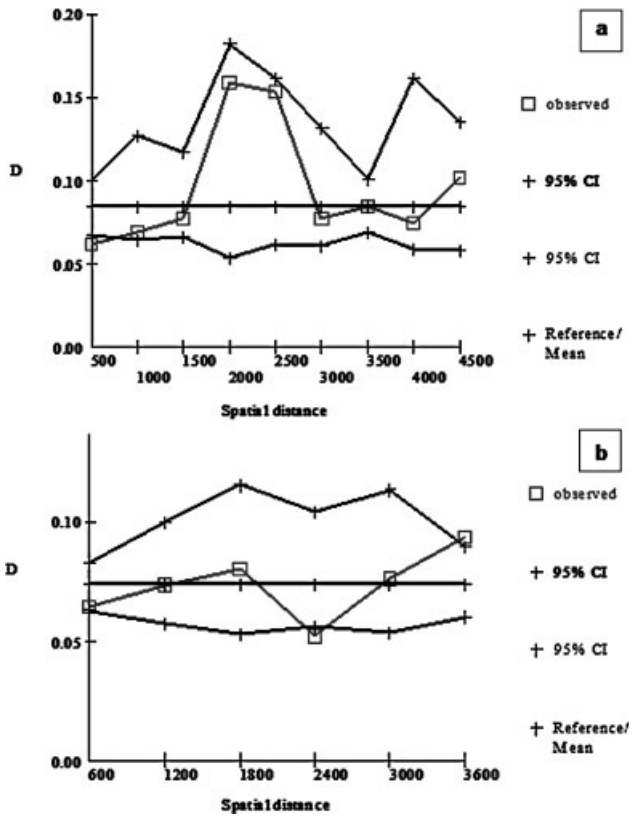

Fig. 5. Distograms of spatial genetic analysis for a) 39 Niger-Congo speaking populations, and b) 31 Bantu speaking populations. Y-axis = Nei genetic distances; X-axis = distance classes (km).

tions were grouped according to geographical and linguistic criteria (Table 3). Analyses were first conducted for 39 populations belonging to the Niger-Congo linguistic family (1,500 individuals). The F_{ST} value calculated for the entire Niger-Congo sample indicates that 5.1% of the genetic variation is due to interpopulation differences. When the populations were partitioned into geographical or linguistic groups we obtained similar values of F_{CT} and F_{ST}, while the F_{SC} value (i.e. the apportionment of variation among populations within groups) was slightly lower in the geographical grouping than in the linguistic one. The overall F_{ST} value calculated for the entire Bantu-speaker sample (mtDNA HVSI: 31 populations, 1,102 individuals) is low, albeit significant, indicating low genetic differentiation between populations (Table 3). To assess the presence of a geographical or linguistic structure within the Bantu language sub-family, we performed a second set of analyses grouping the populations according to linguistic and geographical subgroup criteria. The *F*-statistic values are very similar in the analyses performed using different groups (Table 3). Nevertheless, geographical criteria grouping with a slight higher statistical support was obtained when western Bantu-speaking populations were pooled in a single group. Anyway, no differences in the genetic homogeneity index between the two groups were pointed out when we performed an AMOVA analysis on western and eastern population separately (F_{ST} values = 0.0175 and 0.0237, respectively for western and eastern groups).

To verify the existence of gene flow between Bantu-speaking populations and between Bantu and non-Bantu speakers, we estimated migration rates using Migrate analysis. The estimates averaged over three different runs are reported in Table 4. High level of migration between all Bantu-speaking populations can be pointed out. Moreover, a relevant gene flow between eastern Bantu speakers and other eastern non-Bantu groups is evident.

DISCUSSION

The aim of the present study is to clarify some aspects of the Bantu-speaker recent migration event through the analysis of two previously unstudied Bantu-speaking populations from eastern and southern Africa, namely the Hutu from Rwanda and the Shona from Zimbabwe, in conjunction with previously published data from 37 Niger-Congo speaking populations. Archeological and linguistic data suggest that both groups arrived in their current settlement areas during the Bantu-speaker expansion. Indeed, we found in the two populations almost all the mtDNA haplogroups previously indicated as Bantu-speaker markers (Bandelt et al., 2001; Salas et al., 2002; Beleza et al., 2005). However, there are some differences in haplogroup composition between the two populations. The haplogroup profile that characterize the Shona from Zimbabwe seems to indicate a higher contribution of the western Bantu speaking populations to the examined Shona group than previously observed in other southeastern Bantu-speaking populations (Pereira et al., 2001; Salas et al., 2002), as represented in the CA graph. This hypothesis is supported by migration rate estimates that indicate a number of exchanged migrants between Shona and southwestern Bantu speakers that is twice that found between other southeastern and southwestern populations. This pattern could be explained by the geographical position of our sample, in the northern internal region of Zimbabwe and thus more exposed to contacts with people coming from the western regions, but also by a migratory pattern different from that of other southeastern Bantu-speaking populations. Archeological data from different Zimbabwean and Zambian sites suggest that settlement of northwestern Mashonaland during the Early Iron Age can be ascribed to people coming from the western stream of Bantu-speaker migrations (Phillipson, 1976; Bisson, 1992). Moreover, the name 'Shona' encompasses several groups of people in Zimbabwe and western Mozambique speaking different dialects and with a long history of dispersion throughout the region. It is not surprising then to find genetic differences between different Shona groups that could reflect different patterns of gene flow from neighboring populations or different settlement histories. In addition, some gene flow from the preexisting southern Khoisan populations is suggested by the presence of haplogroups L0k1 and L0d (Salas et al., 2002) and confirmed by migration rate estimates. On the other hand, the presence in the Hutu population of haplogroups (L0f, L3x, L4g, L5, M) typical of other non-Bantu eastern African groups (Kivisild et al., 2004; Tishkoff et al., 2007; Castrì et al., 2008) points to the occurrence of substantial gene flow from populations of eastern African origins, most probably the pastoral Tutsi populations. Again, this result is confirmed by migration rates estimates, with more than 30 exchanged migrants per generation.

TABLE 2. Diversity and neutrality indices in Niger-Congo populations

Geographic and linguistic group	*n*	No. haplotypes	H[a] (SD)	Nucleotide diversity (SD)	MNPD[b] (SD)	Tajima's D (*P*)	Fu's *F*s (*P*)
Niger Congo non Bantu	349	180	0.989 (0.002)	0.024 (0.012)	7.923 (3.693)	−1.606 (0.040)	−24.408 (0.000)
Western Bantu	407	187	0.990 (0.001)	0.029 (0.014)	10.163 (4.652)	−1.375 (0.074)	−24.044 (0.001)
Eastern Bantu	88	66	0.991 (0.004)	0.031 (0.016)	11.091 (5.087)	−1.199 (0.115)	−24.559 (0.000)
South Western Bantu	182	107	0.978 (0.006)	0.026 (0.013)	8.992 (4.161)	−1.411 (0.070)	−24.522 (0.000)
South Eastern Bantu	473	177	0.974 (0.003)	0.026 (0.013)	9.443 (4.342)	−1.105 (0.134)	−24.076 (0.002)

[a] Haplotype diversity (Standard Deviation).
[b] Mean number of pairwise sequences (Standard Deviation).

TABLE 3. AMOVA analysis of mtDNA data from Niger Congo and Bantu populations

Grouping criteria	Within pop % variation F_{ST}	Among pop % variation F_{SC}	Among groups % variation F_{CT}
39 Niger-Congo populations			
1 group	94.91, 0.051		
6 linguistic groups (Mande, Atlantic, Adamawa, Defoid, Narrow Grassfields, Narrow Bantu)	93.27, 0.067	3.56, 0.037	3.17, 0.032
5 geographic groups (W, WC, SW, E, SE)	94.11, 0.059	2.22, 0.023	3.66, 0.037
31 Narrow Bantu speaking populations			
1 group	96.31, 0.037		
2 linguistic groups (Northwest, Central)	95.69, 0.043	3.12, 0.032	1.19, 0.012
10 linguistic groups (A, B, E, J, F, H, R, S, P, N)	96.13, 0.039	1.27, 0.013	2.60, 0.026
2 geographical groups (W, E)	95.33, 0.047	2.53, 0.026	2.14, 0.021
3 geographical groups (W, E, SE)	95.32, 0.047	2.12, 0.022	2.56, 0.026
3 geographical groups (W, SW, E)	95.73, 0.043	2.61, 0.027	1.66, 0.017
4 geographical groups (WC, SW, E, SE)	95.72, 0.043	2.16, 0.022	2.12, 0.021

P-value < 0.000001 for all indices.

TABLE 4. Migrate analysis on some African and Bantu-speaking populations

Population	Shona	Hutu	SEBantu	EBantu	EAfrica	CWBantu	SWBantu	KhoiSan
Shona	–							
Hutu	10,77	–						
SEBantu	10,75	1,93	–					
EBantu	14,14	42,87	**16,08**	–				
EAfrica	11,54	**30,20**	5,32	**42,91**	–			
CWBantu	15,78	8,48	13,27	33,33	4,91	–		
SWBantu	**15,60**	2,29	**7,01**	34,64	4,74	31,78	–	
KhoiSan	1,98	0,97	1,98	28,20	13,63	1,64	1,54	–

Short chains = 10 (used trees 10,000/200,000); long chains = 3 (used trees 100,000/20,000,000); average values on three independent multiple runs.

Our analysis of mtDNA variability of the Shona and the Hutu, in conjunction with previously published data of other Bantu-speaking groups, also helped us to clarify some aspects of the migration events that involved Bantu farmers in sub-Saharan Africa. AMOVA showed that neither linguistic nor geographical structures are strongly supported by the data. On the other hand, we detected high level of gene flow between all Bantu-speaking populations. Furthermore, estimates of migration rates between Bantu-speaking and other African populations suggest the occurrence of varying degrees of gene admixture with populations encountered en-route to southern Africa. This high level of gene flow is evident not only in the Hutu and the Shona, but also in other Bantu-speaking populations, and it is clear also in the spatial structure revealed through SGS analysis. All these results suggest the existence, after a first expansion of Bantu-speakers (or proto-Bantu-speakers and associated cultures), of extensive interactions and trading within Bantu-speaking populations and with other non-Bantu speaking populations. A massive movement of people with conquest and displacement of preexisting populations should have resulted in a clear genetic cline as expected under the IBD model. Hence, our results fit better to the model developed for linguistic data by Ehret (2001) suggesting a gradual spread of Bantu-speakers, with strong interactions between the different lines of Bantu-speaker descent. Indeed, the continuous or tree model supporting a single "great" expansion of farmers from western-central Africa should have led to high divergence between distant populations (Holden, 2002). Our results are also in agreement with recent archeological findings, that suggest the existence of long-range trade routes connecting ancient Bantu-speaking communities between them and with other populations (Chami, 2001). Unfortunately, the lack of data on Bantu-speaking populations from central Africa (C.A.R. and Dem. Rep. Congo) does not allow us to determine from a genetic point of view whether there was an early expansion from western-central Africa or whether the migra-

tion stream started after an initial settlement in the internal rainforest region.

In conclusion, our data emphasize the role that population admixture has played at different times and to varying degrees in the dispersal of Bantu-speaking languages. Very recently, Quintana-Murci et al. (2008) reported a high level of asymmetric gene flow between Bantu-speaking farmers and Pygmies from western-central Africa, while eastern Pygmies showed a haplogroup pattern more similar to that of eastern African populations. It should be stressed that mtDNA data give us indications on demographic and evolutionary processes involving only the maternal lines of descent. Data on the Y-chromosome in Bantu-speaking populations indicate a different pattern of genetic variability, with lower diversity in the Y-chromosome than in mtDNA (Pereira et al., 2002; Beleza et al., 2005) and no evidence of demographic expansion (Pilkington et al., 2008). These results are not surprising and might reflect the strong influence of sociocultural factors on demographic processes accompanying human migrations. In particular, it has been hypothesized that, during agricultural expansion, contacts between food-producers and populations they encountered along the way were characterized by unidirectional marriages between hunter/gatherer and/or forager females and farmer males (Destro-Bisol et al., 2004; Wen et al., 2004; Wood et al., 2005; Pilkington et al., 2008). Our data, indicating the assimilation into the expanding farmer groups of mtDNA lineages from neighboring populations, support such a demographic scenario.

ACKNOWLEDGMENTS

The authors are grateful to all the Shona and Hutu people who participated in this project. The authors thank Francesco Olivieri and Dr. Maria Elena Pesaresi (Luisa Guidotti Hospital) for collecting the Shona oral swabs, Marcello Franceschi for helping us in Hutu sampling, Lorena Madrigal and Guido Barbujani for useful comments and suggestions on an early version of the manuscript, and Maja Dembic and Stefania Bertoncini for technical assistance in the laboratory analyses. The authors would also thank four anonymous reviewers for their helpful comments.

LITERATURE CITED

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147.

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.

Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. PNAS USA 98:4563–4568.

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S, Genographic Consortium. 2008. The dawn of human matrilineal diversity. Am J Hum Genet 82:1130–1140.

Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A. 2005. The genetic legacy of western Bantu migrations. Hum Genet 117:366–75.

Bertorelle G, Barbujani G. 1995. Analysis of DNA diversity by spatial autocorrelation. Genetics 140:811–819.

Bisson MS. 1992. A survey of Late Stone Age and Iron Age Sites at Luano, Zambia. World Archaeol 24:234–248.

Bortolini MC, Da Silva WA JR, Zago MA, Elion J, Krishnamoorthy R, Gonçalves VF, Pena SDJ. 2004. The phylogeography of mitochondrial DNA haplogroup L3g in Africa and the Atlantic Slave Trade. Am J Hum Genet 75:523–524.

Castrì L, Garagnani P, Useli A, Laino M, Flamigni E, Pettener D, Donata Luiselli. 2008. Kenyan crossroads: gene flow and migration patterns in six ethnic groups from Kenya. J Anthropol Sci 86:189–192.

Chami FA. 2001. A Response to Christopher Ehret's "Bantu Expansions". Int J Afr Hist Stud 34:647–651.

Degen B, Petit R, Kremer A. 2001. SGS - Spatial Genetic Software: A computer program for analysis of spatial genetic and phenotypic structures of individuals and populations. J Heredity 92:447–448.

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

Ehret C. 1981. The demographic implications of linguistic change and language shift. In: Fyfe C, McMaster D, editors. African historical demography , Vol.2. Edinburgh: University of Edinburgh, Centre of African Studies. p 153–182.

Ehret C. 2001. Bantu expansion: re-envisioning a central problem of early African history. Int J Afr Hist Stud 34:5–41.

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.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47–50.

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.

Holden CJ. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. Proc R Soc Lond B 269:793–799.

Huffman TN. 1982. Archaeology and Ethnohistory of the African Iron Age. Ann Rev Anthropol 11:133–150.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75:752–770.

Liesegang G, Seitz S, Winter JC. 1979. Das Aequatoriale Ostafrica. In: Baumann H, editor. Die Volker Afrikas und Ihre traditionellen Kulturen, Teil 2. Wiesbaden: Franz Steiner Verlag GmbH. p 25–33.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249.

Newman JL, 1995. The peopling of Africa: a geographic interpretation. New Haven: Yale University Press.

Pereira L, Gusmao L, Alves C, Amorim A, Prata MJ. 2002. Bantu and European Y-lineages in Sub-Saharan Africa. Ann Hum Genet 66:369–78.

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 65:439–58.

Phillipson DW. 1976. Archaeology and Bantu linguistics. World Archaeol 8:65–82.

Phillipson DW. 1993. African archaeology. Cambridge: Cambridge University Press.

Phillipson DW. 2003. Language and farming dispersal in sub-Saharan Africa, with particular reference to the Bantu-speaking people. In Bellwood P. and Renfrew C, editors. Examining

the farming/language dispersal hypothesis. Cambridge: McDonald Institute, p 170–193.
Pilkington MM, Wilder JA, Mendez FL, Cox MP, Woerner A, Angui T, Kingan S, Mobasher Z, Batini C, Destro-Bisol G, Soodyall H, Strassmann BI, Hammer MF. 2008. Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. Mol Biol Evol 25:517–525.
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.
Quintana-Murci L, Quacha H, Harmanta C, Luca F, Massonnet B, Patin E, Sica L, Mouguiama-Daoudad P, Comas D, Tzur Shay, Balanovsky O, Kidd KK, Kidd JR, van der Veen L, Hombert JM, Gessain A, Verdu P, Froment A, Bahuchet S, Heyer E, Dausset J, Salas A, Behar DM. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. PNAS 105:1596–1601.
Salas A, Richards M, De la Fe T, Lareu M, Sobrino B, Sanchez-Diz P, Macaulay V, and 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. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465.
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.
Sturrock K, Rocha J. 2000. A multidimensional scaling stress evaluation table. Field Methods 12:49–60.
Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB. 2000. Y chromosomes travelling south: the cohen modal haplotype and the origins of the Lemba—the 'Black Jews of Southern Africa'. Am J Hum Genet 66:674–686.
Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernandopulle N, Lema G, Nyambo TB, Ramakrishnan U, Reed FA, Mountain JL. 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. Mol Biol Evol 24: 2180–2195.
Tofanelli S, Boschi I, Bertoneri S, Coia V, Taglioli L, Franceschi MG, Destro-Bisol G, Pascali V, Paoli G. 2003. Variation at 16 STR loci in Rwandans (Hutu) and implications on profile frequency estimation in Bantu-speakers. Int J Legal Med 117:121–126.
Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339–345.
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.
Vansina J. 1995. New linguistic evidence and the Bantu expansion. J Afr Hist 36:173–195.
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507.
Wen B, Xie XH, Gao S, Li H, Shi H, Song XF, Qian TZ, Xiao CJ, Jin JZ, Su B, Lu D, Chakraborty R, Jin L. 2004. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. Am J Hum Genet 74:856–865.
Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur J Hum Genet 13:867–876.