

# Genetic and Demographic Implications of the Bantu Expansion: Insights from Human Paternal Lineages

Gemma Berniell-Lee,\* Francesc Calafell,\*† Elena Bosch,\*† Evelyne Heyer,‡ Lucas Sica,§ Patrick Mougouma-Daouda,|| Lolke van der Veen,¶ Jean-Marie Hombert,¶ Lluís Quintana-Murci,# and David Comas\*†¶

\*Institute of Evolutionary Biology (CSIC-UPF), CEXS-UPF-PRBB, Barcelona, Catalonia, Spain; †CIBER de Epidemiología y Salud Pública, Spain; ‡Muséum National d'Histoire Naturelle—Centre National de la Recherche Scientifique-P7 Unité Mixte de Recherche 5145, Eco-Anthropologie, Musée de l'Homme, Paris, France; §Centre International de Recherches Médicales de Franceville, Franceville, Gabon; ||Université Omar Bongo, Libreville, Gabon; ¶Laboratoire Dynamique du Langage, Unité Mixte de Recherche 5596, Institut des Sciences de l'Homme, Lyon, France; and #Human Evolutionary Genetics Unit, Centre National de la Recherche Scientifique URA3012, Institut Pasteur, Paris, France

The expansion of Bantu languages, which started around 5,000 years before present in west/central Africa and spread all throughout sub-Saharan Africa, may represent one of the major and most rapid demographic movements in the history of the human species. Although the genetic footprints of this expansion have been unmasked through the analyses of the maternally inherited mitochondrial DNA lineages, information on the genetic impact of this massive movement and on the genetic composition of pre-Bantu populations is still scarce.

Here, we analyze an extensive collection of Y-chromosome markers—41 single nucleotide polymorphisms and 18 short tandem repeats—in 883 individuals from 22 Bantu-speaking agriculturalist populations and 3 Pygmy hunter-gatherer populations from Gabon and Cameroon. Our data reveal a recent origin for most paternal lineages in west Central African populations most likely resulting from the expansion of Bantu-speaking farmers that erased the more ancient Y-chromosome diversity found in this area. However, some traces of ancient paternal lineages are observed in these populations, mainly among hunter-gatherers. These results are at odds with those obtained from mtDNA analyses, where high frequencies of ancient maternal lineages are observed, and substantial maternal gene flow from hunter-gatherers to Bantu farmers has been suggested. These differences are most likely explained by sociocultural factors such as patrilocality. We also find the intriguing presence of paternal lineages belonging to Eurasian haplogroup R1b1\*, which might represent footprints of demographic expansions in central Africa not directly related to the Bantu expansion.

## Introduction

Africa has been the theater of multiple human movements—both within and out of Africa—throughout its history that have shaped its genetic landscape. Undoubtedly, the Bantu expansion is one of them. The term “Bantu expansion” refers to a very complex phenomenon involving the transmission of culture, language, and technology, and most likely, of genes. Having been described as the greatest population movement in recent African history (Diamond 1997), it is thought that ~5,000 years before present, Bantu-speaking people left their homeland in north-western Cameroon/southern Nigeria and spread throughout sub-Saharan Africa. During the expansion, which was probably related to the spread of agriculture, and to some extent, to the emergence of iron technologies (Phillipson 1993; Newman 1995; Vansina 1995), the new coming Bantu-speaking people led to the isolation or to the admixture with neighboring hunter-gatherer populations, namely, Pygmies and Khoisans (Cavalli-Sforza 1986). The result was the simultaneous spread of Bantu cultures, languages, and genes across sub-equatorial Africa (Diamond and Bellwood 2003), where nowadays most sub-Saharan Africans are speakers of Bantu languages. Given that the expansion did not follow a single continuous migration route, but rather, that it involved at least two major dispersals with different expansion centers (one in the west and one in the east) (Oslisly 1995), different geographical constraints, and at different times, it is

not surprising that differences in the genetic composition of the different Bantu areas have been found, especially in terms of the degree of assimilation of hunter-gatherer populations (Thomas et al. 2000; Pereira et al. 2001, 2002; Salas et al. 2002; Plaza et al. 2004; Beleza et al. 2005).

Although the demographic intensity and impact of the Bantu expansion still remains unclear (Bakel 1981; Phillipson 1993), its effects are visible from the current genetic diversity patterns observed in sub-Equatorial Africa. Traditionally, mitochondrial DNA (mtDNA) has been widely used to measure the genetic impact of the Bantu expansion: some mtDNA lineages (such as L0a, L2a, L3b, and L3e) have been postulated as genetic footprints of the Bantu expansion (Pereira et al. 2001; Salas et al. 2002, 2004; Plaza et al. 2004; Beleza et al. 2005; Wood et al. 2005) given their frequency and diversity patterns in Bantu- and non-Bantu-speaking populations. Some lineages of more ancient origin (different clades within the deep-rooting haplogroup L1c) might be ancient remnants of the diversity present in West/Central Africa before the Bantu expansion (Batini et al. 2007; Quintana-Murci et al. 2008). Albeit (particularly in comparison with mtDNA) variation at the Y-chromosome has not yet been thoroughly investigated in Africa, a number of Y-chromosome haplogroups, such as E1b1a (previously named E3a), E2, and B2a, have been proposed as paternal signatures of the Bantu expansion (Underhill et al. 2000; Cruciani et al. 2002; Beleza et al. 2005). Other paternal lineages (such as B2b or different clades within A) have been mainly observed among hunter-gatherers and suggested to represent ancient remnants of the paternal diversity in sub-Saharan Africa prior to the Bantu expansion (Underhill et al. 2000; Cruciani et al. 2002). In addition, lineages belonging to Eurasian

Key words: human population genetics, genetic diversity, Y-chromosome phylogeny, Bantu expansion.

E-mail: david.comas@upf.edu.

*Mol. Biol. Evol.* 26(7):1581–1589. 2009

doi:10.1093/molbev/msp069

Advance Access publication April 15, 2009

haplogroup R have been found in northern Cameroon and have been claimed to result from back migrations from Eurasia into Africa (Cruciani et al. 2002).

Although the genetic evidence is poor, it has been suggested that the Bantu expansion influenced the Y-chromosome gene pool of sub-Saharan Africans to a greater extent than the mtDNA pool due to sex-biased rates of admixture between Bantu farmers and local hunter-gatherers (Destro-Bisol et al. 2004; Wood et al. 2005). For instance, sociocultural taboos have prevented maternal Bantu-to-Pygmy and paternal Pygmy-to-Bantu gene flow (Cavalli-Sforza 1986; Destro-Bisol et al. 2004).

A geographically detailed coverage, coupled with ethnologically well-defined populations in a given geographic area, is essential to infer past demographic events with precision. Here, we focus on what was presumably one of the first stepping stones of the Bantu expansion: west Central Africa, a region where Pygmy populations have retained their cultural and biological identity and coexist with non-Pygmy Bantu speakers. In this context, genetic data from west central African populations are scarce. We have analyzed the paternal lineages of >800 individuals in this region with three main aims: 1) determine the genetic structure of Pygmy populations; 2) measure the extent and symmetry of the gene flow between Pygmies and non-Pygmies; and 3) trace the spread of the presumably non-African R1b1\* haplogroup in order to refine the hypotheses to explain its presence in Africa.

## Materials and Methods

### Population Sampling

We obtained blood samples from a total of 883 unrelated healthy males from 21 populations from Gabon and four populations from Cameroon (fig. 1). Gabonese samples comprise 20 Bantu-speaking agriculturalist populations and one Pygmy population (Baka). Cameroonian samples comprise two Bantu-speaking agriculturalist populations (Fang and Ngumba) and two Pygmy populations (Baka and Bakola). All individuals were interviewed by linguists and anthropologists in order to verify their ethnic affiliations and an informed consent was obtained for each individual. DNA extraction was carried out using a standard phenol chloroform method.

### Y-Chromosome Typing

All the individuals were typed for 35 Y-single nucleotide polymorphisms (SNPs) in a single reaction using a multiplex assay system for major haplogroup screening with SNPlex technology (Applied Biosystems, Foster City, CA) (Berniell-Lee et al. 2007). In order to refine the phylogenetic resolution in some Y-chromosome branches, some individuals were further typed for six SNPs. Markers M150, P25, M17, M18, and M269 were typed using SNaPhot technology (Applied Biosystems) and M73 was typed using TaqMan technology (Applied Biosystems). Haplogroup names follow the nomenclature recently proposed (Karafet et al. 2008). Eighteen highly informative Y-short tandem repeats (STRs) were typed in the form of three previously de-

scribed multiplex reactions: MSI and EBF (Bosch et al. 2002), and CTS (Ayub et al. 2000). Polymerase chain reactions (PCRs) were performed in a 10- $\mu$ l final volume: AmpliTaqGold Buffer (1 $\times$ ), dNTPs (0.2 mM), AmpliTaqGold MgCl<sub>2</sub> (2 mM), AmpliTaqGold DNA Polymerase (1 unit), multiplex primer mix (1 $\times$ ), and DNA (5–10 ng). PCR cycling conditions for multiplex MSI were modified from Bosch et al. (2002) as follows: 95 °C for 10 min; 10 cycles of 95 °C for 1 min, 60–55 °C for 1.5 min (–0.5 °C per cycle), and 72 °C for 1 min; 20 cycles of 95 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min; final extension of 72 °C for 10 min. PCR cycling conditions for multiplexes CTS and EBF were as described by Ayub et al. (2000) and Bosch et al. (2002), respectively. Diluted PCR products were mixed with Genescan 400 HD[ROX] size standard (Applied Biosystems) and run on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). The amplicon size was analyzed using GeneMapper Software v3.7 (Applied Biosystems). Allele designations are in accordance with the Y Chromosome Haplotype Reference nomenclature (<http://www.yhrd.com>). SNP and STR data are available in the supplementary table 1, Supplementary Material online.

### Statistical Analyses

For each of the 25 populations, diversity measures were computed both for STR haplotypes and SNP haplogroups using Arlequin v2.0 software (Schneider et al. 2000). The total number of haplotypes and the number of haplotypes shared between populations were determined by a simple counting scheme. The frequency of the Y-chromosome haplogroups found, together with the modal Bantu haplotype described by Thomas et al. (2000) and of its one-step neighbors described by Pereira et al. (2002) were calculated by a simple counting scheme. Population genetic structure was tested through analysis of molecular variance (AMOVA) using the Arlequin program v2.0 (Schneider et al. 2000).

Dating was performed using the method described by Goldstein et al. (1996) who established that the variance in repeat size in STRs is proportional to expansion times and mutation rates. This method offers the advantage of not requiring the establishment of the precise phylogenetic relationships among all haplotypes, which, given the present sample size, may be difficult. Expansion times and their standard deviations (SDs) were calculated using 11 STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, and DYS460), whose mutation rates have been individually estimated (Gusmao et al. 2005). The allelic variance of each STR was divided by the estimated mutation rate, and the mean of the variances was multiplied by 25 (intergeneration time in years).

Genetic relationships between the haplotypes within specific haplogroups were analyzed using the Network program v 4.1.1.1, applying the median-joining method (Bandelt et al. 1995). STR weighting was applied according to the molecular variance of each marker. Higher weights were given to the least variable loci, with weight being inversely proportional to variance. Multi-copy marker DYS385 was considered as two separate loci. The nonperfect repeat 13.2 for this locus, found only in haplogroup

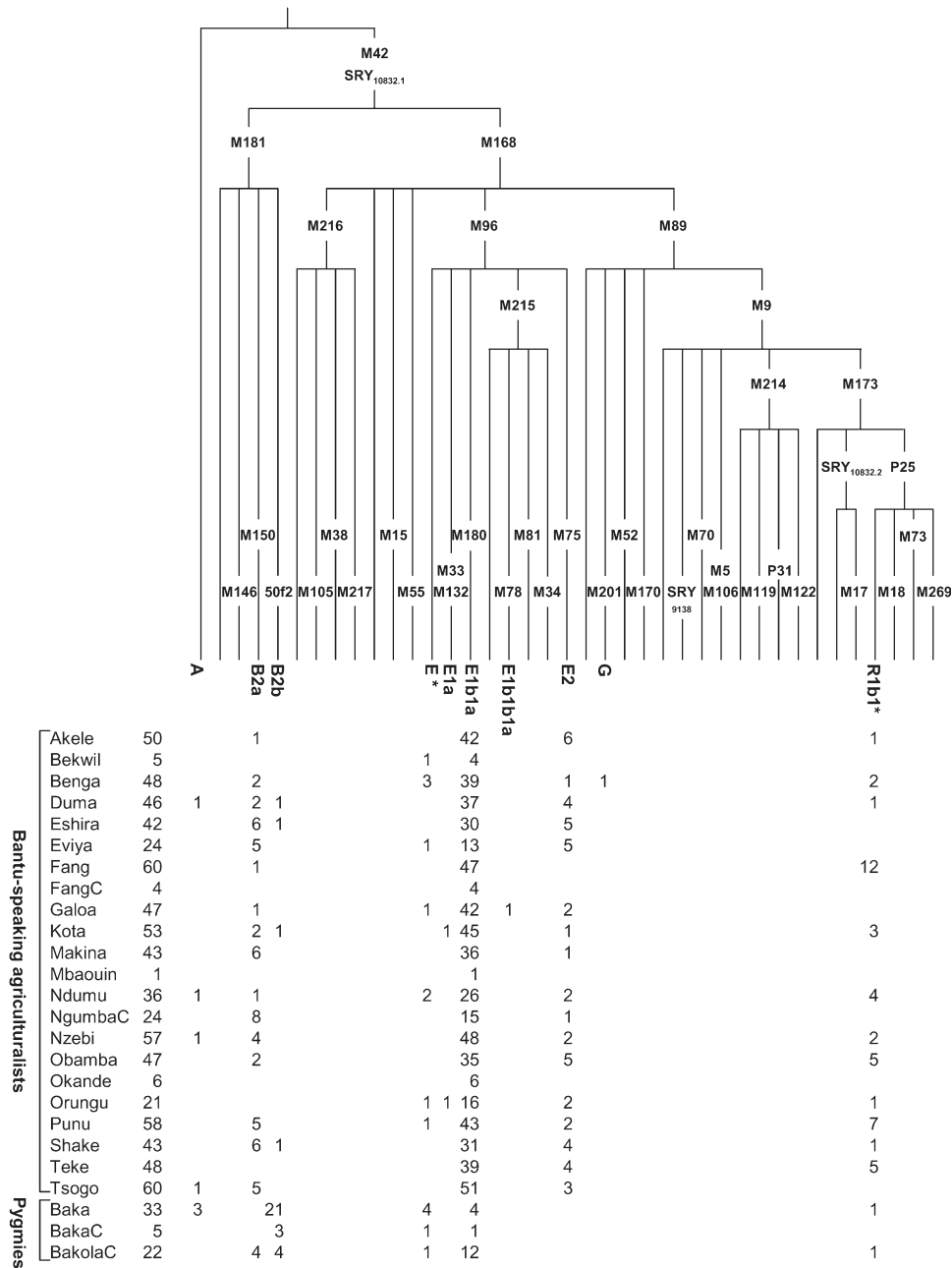


FIG. 1.—Absolute Y-chromosome haplogroup frequencies found in west central African samples. Samples from Cameroon are shown by a final capital C, whereas the rest of the samples were from Gabon. The 41 SNPs genotyped are shown in the branches of the Y-chromosome phylogeny and only haplogroup names found in the sample set are shown (Karafet et al. 2008).

R1b1\*, was incorporated into the analysis by coding it as an additional locus with two states, 1 for the haplotypes carrying perfect alleles and 2 for the imperfect 13.2 alleles. Given that the recurrent generation of this allele is highly unlikely, and could be considered to be equivalent to a unique event polymorphism, it was given the maximum weight when constructing the R1b1\* network.

In order to establish the genetic relationships between west Central African samples and the rest of the sub-Saharan continent, a Correspondence Analysis (CA) for haplogroup frequencies was calculated using STATISTICA 6.0 (Stat Soft, Inc). A total of 74 African populations (supplementary table 2, Supplementary Material online) were used in the CA:

25 populations from the present study and 49 populations belonging to different linguistic groups (Afro-Asiatic, Nilo-Saharan, Khoisan, and Niger-Congo). Given that the phylogenetic resolution was not the same for all the studies used for comparison, adjustments have been made in order to obtain similar/equivalent levels of haplogroup discrimination, defining in this way a total of 22 haplogroups.

**Results**

**Y-Chromosome Lineages in West Central Africa**

The 883 male samples analyzed from west Central Africa were classified into 10 different haplogroups according

to the recently published Y-Chromosome Phylogeny (Karafet et al. 2008) (fig. 1). The haplogroup diversity of these Y-lineages was low ( $0.414 \pm 0.020$ ) with respect to that reported for other areas of Africa (Beleza et al. 2005). However, when hunter-gatherer Pygmies and Bantu agriculturalists were considered separately, the haplogroup diversity was much higher for the former ( $0.694 \pm 0.044$ ) than for the latter ( $0.362 \pm 0.021$ ).

On the whole, most of the samples belonged to previously described African lineages especially common in sub-Saharan Africa. Specifically, most of these lineages have been associated either with Bantu-speaking people—E1b1a (E3a according to The Y Chromosome Consortium 2002), B2a, and E2—or to Pygmy populations (haplogroup B2b). We also observed traces of haplogroups A, E\*, E1a, and E1b1b1a (E3b1 according to The Y Chromosome Consortium 2002), which are found at low frequencies across the African continent (Underhill et al. 2000, 2001; Cruciani et al. 2002; Wood et al. 2005). Interestingly, almost 5% of the individuals here analyzed belonged to Eurasian haplogroup R1b1\*.

Haplogroup E1b1a, previously proposed as being a marker of the Bantu expansion (Passarino et al. 1998; Scozzari et al. 1999; Underhill et al. 2001; Wood et al. 2005), was the most frequent haplogroup in our sample set (76%), reaching a frequency of almost 80% in agriculturalists and 28% in Pygmy samples (fig. 1). Notably, it was also the most frequent haplogroup in the Bakola Pygmy group from Cameroon (55%). When the internal STR diversity of haplogroup E1b1a was considered, Bantu-speaking farmers presented higher diversity ( $6.56 \pm 3.11$  pairwise differences) than Pygmies ( $5.89 \pm 2.96$ ), and a clear starlike shape was observed in the network of haplotypes (supplementary fig. 1, Supplementary Material online). The expansion date of the E1b1a haplogroup was estimated at 5,800 years (SD 7,200), in agreement with the expansion of Bantu languages. It is worth mentioning that the 14 different E1b1a STR haplotypes found in Pygmies were quite divergent and did not form a cluster in the network, being found at the tips of the tree and with only three of them being shared with Bantu-speaking agriculturalists. The modal Bantu haplotype previously described (Thomas et al. 2000), consisting of alleles 10, 21, 13, 11, and 15 for STRs DYS391, DYS390, DYS393, DYS392, and DYS19, respectively, was found in all Bantu-agriculturalist populations (excluding the Fang from Cameroon and the Mbaouin, with less than five individuals analyzed), at frequencies ranging from 12% to 43%. This haplotype was also observed among Bakola Pygmies (14%) and at lower frequencies among Baka Pygmies from Gabon (6%). High frequencies of one-step neighbors of the founder Bantu haplotype (haplotypes at one mutational step from the modal Bantu haplotype) (Pereira et al. 2002) were also observed in 18 of the 20 Bantu-agriculturalist populations.

Haplogroup B2a, which has been previously associated with Bantu agriculturalists (Beleza et al. 2005), was the second most frequent haplogroup (7%) in the west Central African sample set. Haplogroup B2a was found in 16 of the Bantu-agriculturalist populations and in four Bakola individuals. The estimated expansion time for the B2a network was 5,200 years (SD 5,300) (supplementary fig. 2,

Supplementary Material online), again in agreement with the hypothesized demographic expansion of Bantu agriculturalists.

Haplogroup B2b was present in all the Pygmy populations of the sample set, showing very high frequencies in both Baka populations (over 60%), and lower frequencies in the Bakola population (around 27%). However, this haplogroup was virtually absent among agriculturalist populations, only being observed in four individuals (0.4%) (fig. 1). The estimated expansion time for the B2b network was 11,000 years (SD 9.0), with haplotypes found in Bantu agriculturalists lying at the tips of the tree (supplementary fig. 3, Supplementary Material online) and showing lower STR diversity ( $8.83 \pm 5.17$ ) than those found in Pygmies ( $9.84 \pm 4.64$ ).

The third most frequent haplogroup in our sample set was E2 (5.7%) (supplementary fig. 4, Supplementary Material online). Although this lineage is found at low frequencies across the entire African continent (Cruciani et al. 2002), its subbranch E2b has been found to be more frequent among Bantu agriculturalists (Underhill et al. 2000; Cruciani et al. 2002), and it is therefore thought to have been carried by the Bantu expansion (Cruciani et al. 2002).

A remarkable finding of our study is the substantial number of individuals belonging to haplogroup R1b1\* (5.2%). Surprisingly, it has been previously observed in northern Cameroon (40%) at high frequencies (Cruciani et al. 2002) and at lower frequencies in southern Cameroon (1.12%) (Cruciani et al. 2002), Oman (1%), Egypt (2%), and Hutu from Rwanda (1%) (Luis et al. 2004). The presence of this lineage in Africa has been claimed to be a genetic signature of a possible backflow migration from west Asia into Africa (Cruciani et al. 2002). Here we observe R1b1\* in 12 Bantu-agriculturalist populations (ranging from 2% to 20%) and in two Pygmy individuals. A network of R1b1\* haplotypes performed using STR data (fig. 2) shows two main clusters, without any population structure. Interestingly, the estimated expansion time for these haplotypes—7,000 years (SD 8,100)—precedes the time at which the Bantu expansion occurred.

#### Genetic Structure and Population Relationships

An AMOVA based on haplogroup composition was performed in order to assess the level of population structure in west Central Africa. When all samples, Bantu agriculturalists and Pygmy hunter-gatherers, were considered as a single group, some genetic heterogeneity between the samples analyzed was found (10.8%,  $P < 0.0001$ ). When Bantu agriculturalists and Pygmies were analyzed independently, the heterogeneity previously found was partially due to differences between Pygmy populations (19.3%,  $P = 0.001$ ), whereas Bantu agriculturalists were found to be very homogeneous genetically (2.3%,  $P = 0.003$ ). In addition, a large proportion of the total Y-chromosome diversity observed (37.6%,  $P = 0.002$ ) was due to differences between Bantu agriculturalists and Pygmy samples, showing a high genetic difference between these two groups.

In order to establish the genetic relationships between west Central African samples and the rest of the sub-Saharan continent, a CA was performed (fig. 3). The first and second



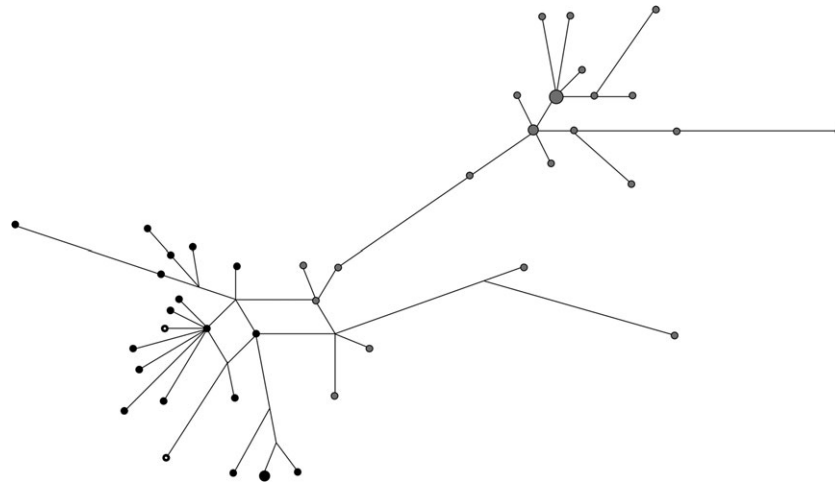


FIG. 2.—Median-joining network of haplogroup R1b1\* based on the data from 18 Y-STRs. Circles are proportional to the individual frequencies. Black circles represent individuals with the imperfect repeat 13.2 allele for DYS385. The only two Pygmy individuals found in the network are represented by a white dot.

dimensions capture 27.7% and 15.2% of the total inertia, respectively, and group the populations according to geographic and linguistic proximities with few exceptions. The first axis mainly separates Afro-Asiatic populations, which are characterized by high frequencies of haplogroup E1b1b (E3b according to The Y Chromosome Consortium 2002) and its derivatives, K, T (K2 according to The Y Chromosome Consortium 2002), and F(xG,I,K) from the rest of populations. However, the scattered distribution of these populations within the Afro-Asiatic cluster indicates a high heterogeneity: 11.29% ( $P < 0.0001$ ) of the genetic variance within Afro-Asiatic samples was due to differences between populations. On the opposite edge of the first axis, the spatial distribution of the Niger-Congo populations (both Bantu and non-Bantu) most likely reflects the high frequency of haplogroups E1b1a and B2a, without any clear appreciable differences between Bantu and non-Bantu speakers (zoomed view in fig. 2): A mere 3.28% ( $P = 0.001$ ) of the genetic variance within the Niger-Congo group is due to differences between Bantu and non-Bantu groups. Compared with non-Bantu Niger-Congo speakers, Bantu populations are more homogeneous in their haplogroup composition: 16.23% ( $P < 0.0001$ ) and 6.92% ( $P < 0.0001$ ) of the genetic variance within groups, respectively. However, a few Bantu populations (e.g., Fang, Southern Cameroon Bantu, Punu, etc.) are scattered in the CA due to the presence of haplogroup R1b1\*. The second axis positions Khoisan, Pygmy, and Nilo-Saharan samples on one edge due to their high frequency of haplogroups B2b and A.

Pygmy samples are clustered in two separate groups; one Pygmy cluster represented by the two western Baka Pygmy samples located close to the Khoisan samples, and another cluster formed by Pygmies from Central Africa (Biaka and Mbuti) together with the Bakola population from Cameroon, located close to Bantu-agriculturalist samples. Differences among all Pygmy samples represent 12.92% ( $P < 0.0001$ ) of the total genetic variance and are not related to the geographical classification of Pygmies into western and eastern groups: No significant differences ( $P = 0.598$ ) in Y-chromosomal haplogroup composition

are found between eastern (Mbuti) and western (Baka, Bakola, and Biaka) samples. The difference in the spatial distribution of the two Pygmy groups most likely reflects the high frequency of haplogroup E1b1a in the latter group of Pygmies, which suggests a stronger Bantu component/influence in these populations than in Baka Pygmies from west Central Africa, who are characterized by high frequencies of haplogroup B2b.

## Discussion

### Pygmy and Bantu Lineages

The general haplogroup profile observed among the Bantu-speaking agriculturalists and Pygmy hunter-gatherer populations of west Central Africa agrees with a genetically homogenizing expansion of the ancestors of Bantu speakers that resulted in an isolation and fragmentation of hunter-gatherer populations. Indeed, Bantu-speaking agriculturalists exhibit a reduced haplogroup diversity compared with hunter-gatherer Pygmies.

Up to 85% of the gene pool of Bantu-speaking agriculturalists belongs to two single lineages: haplogroups E1b1a (~80%) and B2a (~5%), which have previously been related to the Bantu expansion (Underhill et al. 2000). These lineages are highly predominant in west Central African Bantu-agriculturalist groups and could be considered the counterparts of some of the maternal mtDNA subclades within the L0a, L2, and L3 lineages. However, the Y-linked E1b1a and B2a lineages are not only frequent in Bantu-agriculturalist populations, as shown in the CA presented. Most Niger-Congo non-Bantu populations also present high frequencies of these haplogroups that spread during the expansion of Bantu languages, an observation that supports the common origin of populations speaking languages belonging to the major Niger-Congo language family. It has been hypothesized that E1b1a, including its subbranch E1b1a7 (defined by M191, and not tested in the present study), arose in west Central Africa and was later taken southward through a demic expansion (Cruciani et al. 2002). Despite the correlation between STRs and

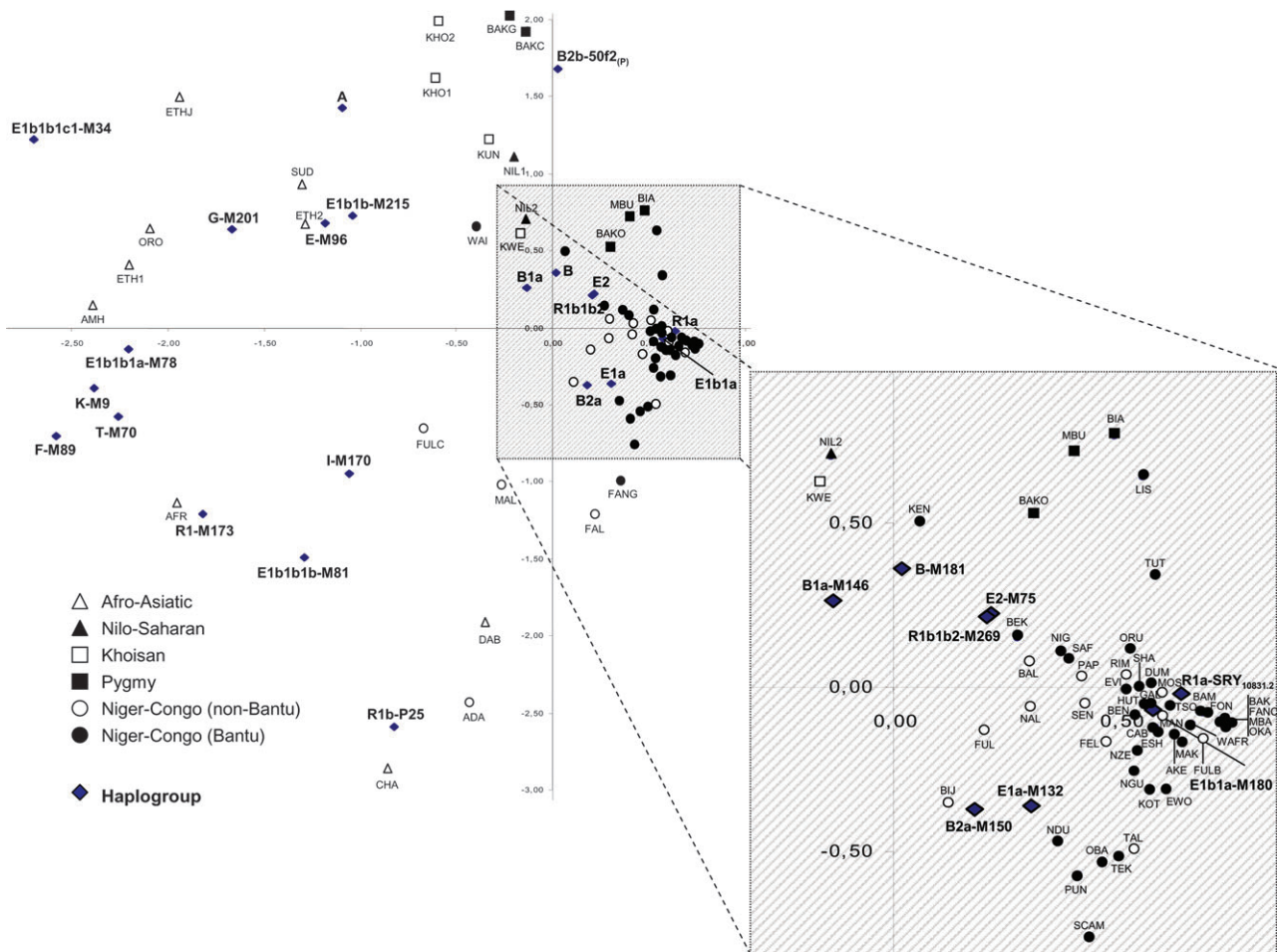


FIG. 3.—CA based on Y-chromosome haplogroup frequencies. Diamonds represent Y-chromosome haplogroups named after Karafet et al. (2008) and show the last derived mutation typed within the Y-chromosome phylogeny. Triangles, squares, and circles represent population samples classified by linguistic/ethnic groups: Afro-Asiatic, Nilo-Saharan, Khoisan, Pygmy, and Niger-Congo (Bantu and non-Bantu). Population abbreviations are as shown in supplementary table 2, Supplementary Material online.

haplogroups in the Y-chromosome (Bosch et al. 1999), the sub-haplogroup lineages within E1b1a were not identifiable in the 18-STR network of haplotypes in the present sample set (supplementary fig. 1, Supplementary Material online). Within B2a, lineages belonging to the B2a1a branch are thought to have originated in west Africa, carried to the east by Bantu farmers, assimilated by central Africans, and subsequently carried southward (Beleza et al. 2005).

In contrast, Pygmies present higher haplogroup diversity with respect to Bantu farmers, which is accounted for by both the presence of pre-Bantu haplogroups (such as B2b and A) and the additional lineages entering the Pygmy gene pool via gene flow from Bantu agriculturalists. Haplogroup B2b has been suggested to have originated early during the history of modern humans (Cruciani et al. 2002; Knight et al. 2003), being especially frequent among Khoisan populations from South Africa (Underhill et al. 2001) and Pygmy populations from the Central African forests (Knight et al. 2003). More specifically, the highest frequencies of B2b are observed among those populations that have experienced little gene flow from Bantu farmers, that is, the Ju|'hoansi (Knight et al. 2003).

### Genetic Structure of Pygmy Populations

The presence of common paternal lineages in eastern and western Pygmies is in strong contrast with the completely nonoverlapping maternal gene pool of these populations. Eastern Pygmies present L0a2, L2a, and L5 mtDNA lineages and do not present L1c lineages, which are found in over 90% of western Pygmies (Quintana-Murci et al. 2008). In contrast, all Pygmy samples here analyzed present paternal B2b frequencies over 25%, reaching 60% in the western Baka Pygmies. Haplogroup B2b has also been observed in Khoisan samples, with some traces in other sub-Saharan populations, pointing toward an ancient origin for this lineage (Underhill et al. 2001). The presence of B2b among Pygmies suggests a common paternal origin for eastern and western Pygmies, being at odds with the apparent lack of maternal ancestry depicted by mtDNA analyses (Quintana-Murci et al. 2008). Sexually asymmetric demographic factors could explain this discrepancy observed in maternal and paternal lineages in Pygmies.

In addition, the presence of haplogroup E1b1a is very heterogeneous in Pygmy populations, being very frequent in Biaka, Mbuti, and Bakola Pygmies (over 40%) and less frequent in Baka (<20%).

## Gene Flow between Pygmies and Non-Pygmies

Our data clearly show that the paternal gene flow between west Central Bantu farmers and Pygmies has been asymmetrical. Although paternal Bantu lineages have been introduced into Pygmies, the opposite has been rare. The predominant Bantu haplogroup E1b1a has been found at a frequency over 25% in Pygmies, especially in Bakola (55%), and B2a has also been found in Bakola (18%). By contrast, the predominant Pygmy haplogroup B2b is observed in less than 1% (4 of 823) of Bantu-speaking agriculturalists. This result is consistent with the hypothesis of asymmetrical gene flow between Pygmies and Bantu farmers due to sex-specific demographic factors (Cavalli-Sforza 1986; Destro-Bisol et al. 2004), with maternal Pygmy-to-Bantu and paternal Bantu-to-Pygmy flow being the rule. The introduction of paternal Bantu lineages into the Pygmy populations could have taken place through extramarital unions between Pygmy females and Bantu-farmer males; Pygmy females being accepted as wives for their great fertility and for their relatively low bride-price, through the adoption of orphans from mixed unions, and through the return of divorced Pygmy women from Bantu-farmer males and their children to Pygmy communities (Destro-Bisol et al. 2004). This is also corroborated by the analyses of female lineages (Batini et al. 2007; Quintana-Murci et al. 2008), where specific Pygmy mtDNA haplogroups, such as L1c1a, have been found in Bantu agriculturalists, whereas the presence of Bantu mtDNA lineages in Pygmy populations has been found to be rare.

Besides this asymmetrical and opposite gene flow unmasked from uniparentally inherited markers, it has been shown that there is a substantial common and deep maternal ancestry between Bantu agriculturalists and Pygmies for the mtDNA, whereas not such a deep ancestry is found for Y-chromosome male lineages. Mitochondrial L1c lineages, prevalent in west Central Africa and found in both groups, have been dated back to 70,000 years (Batini et al. 2007; Quintana-Murci et al. 2008), suggesting a common ancient origin for Bantu agriculturalists and Pygmies. In contrast, the common paternal lineages found in west Central African samples are essentially recent. Only traces of haplogroup A and basal E-M96 are found in both groups, but these haplogroups only account for 5% and 10%, respectively, in Pygmies and 0.5% and 1% in Bantu agriculturalists. This lack of ancient paternal lineages among west Central Africans suggests that the Bantu expansion erased most of the ancient diversity present in the region before the massive demic expansion. In addition, these results clearly indicate that the consequences of the Bantu expansion are more visible from the paternal side than from the maternal side, suggesting that the demic movements associated with the Bantu expansion involved more males than females. Sex-specific gene flow can be directly estimated from the genome regions that unilinearly transmitted; general (and more accurate) estimates of sex-independent admixture can be derived from autosomal markers such as those recently published (Verdu et al. 2009).

## Non-African Lineages in West Central Africa

The analysis of paternal lineages in west Central Africa has shown the predominant presence of haplogroups claimed

to have a Bantu origin (such as E1b1a or B2a), together with a few autochthonous pre-Bantu lineages (such as B2b) (Underhill et al. 2000; Cruciani et al. 2002). Nevertheless, the nonnegligible presence of the R1b1\* lineage in west Central African samples (with a frequency over 5%) might point toward additional demographic expansions within the area besides the “Bantu expansion.” The presence of this haplogroup has also been reported by Cruciani et al. (2002) in Cameroon, as well as in Oman, Egypt, Rwanda (Luis et al. 2004), and Sudan (Hassan et al. 2008), although slightly different genetic markers were analyzed. The presence of this haplogroup in the region is especially puzzling given that, according to the known Y-chromosome phylogeny (Underhill et al. 2000), the geographic origin of the R1b lineage is situated Eurasia and not in Africa. Its sporadic presence, although at low frequencies, in some African populations has been proposed to result from back migrations from Eurasia into Africa during ancient times. The internal STR diversity of this lineage in west Central Africa points toward a putative expansion occurring 7,000 years ago, before the “Bantu expansion.” However, this estimated expansion time could represent an underestimation and may simply show the expansion of a subset of the diversity within haplogroup R1b1\* to west Central Africa through the Bantu expansion, having been shown to be especially frequent in northern Cameroon, near the putative Bantu expansion origin (Cruciani et al. 2002). Surprisingly, no traces of non-African maternal lineages (i.e., mtDNA) have been observed in west Central Africa (Coia et al. 2005; Quintana-Murci et al. 2008), pointing to a putative sexual asymmetrical demographic expansion in Central Africa. Further analyses in an extended group of Central African populations (including Nigeria, Niger, Chad, and Sudan) might be pivotal to shed light on this poorly known demographic event in the region. It is noteworthy that the Fang population is the Bantu-agriculturalist group presenting the highest frequency of R1b1\*. The presence of the Fang in west Central Africa appears to be recent, and they are thought to have entered the region from the north-eastern open grassland plateau during the 17th and 18th centuries (Perrois 2006).

In conclusion, our results demonstrate the recent origin for most paternal lineages in west Central Africa to be the result of the Bantu expansion starting ~5,000 years ago, having erased virtually all previous Y-chromosome diversity of populations inhabiting this region. However, some traces of ancient paternal lineages are observed, mainly among the groups of hunter-gatherers. These results contrast with data drawn from the analyses of maternal lineages in these populations, where ancient and phylogenetically deep lineages are observed and substantial maternal gene flow from hunter-gatherers toward Bantu farmers has been suggested. Finally, the presence of lineages belonging to haplogroup R1b1\* might represent footprints of demographic expansions in Central Africa not directly related to the “Bantu expansion.”

## Supplementary Material

Supplementary tables 1 and 2 and supplementary figures 1–4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).



## Acknowledgments

We thank Mònica Vallés, Anna Pérez-Lezaun, Chiara Batini, Roger Anglada, and Stéphanie Plaza (Universitat Pompeu Fabra, Barcelona) for their technical support and advice. We also thank Alain Froment (Musée de l'Homme, Paris, France) for providing the samples from Cameroon collected during the project ACI-Prosodie "Histoire et diversité génétique des Pygmées d'Afrique Centrale et de leurs voisins." We are also grateful to Chris Tyler-Smith for carefully reading the manuscript and for his valuable advice. We also thank the National Institute of Bioinformatics (www.inab.org), a platform of Genoma España. The research presented was supported by the European Science Foundation EUROCORES Origins of Man, Language, and languages, OMLL program ("Language, culture and genes in Bantu: a multidisciplinary approach to the Bantu-speaking populations of Africa"), the Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain (CGL2007-61016), and Direcció General de Recerca, Generalitat de Catalunya (2005SGR/00608). G.B.-L. received an FI fellowship from the Generalitat de Catalunya.

## Literature Cited

- Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C. 2000. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res.* 28:e8.
- Bakel M. 1981. The "Bantu" expansion: demographic models. *Curr Anth.* 22:688–691.
- Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics.* 141:743–753.
- Batini C, Coia V, Battaglia C, Rocha J, Pilkington MM, Spedini G, Comas D, Destro-Bisol G, Calafell F. 2007. Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. *Mol Phylogenet Evol.* 43:635–644.
- Beleza S, Gusmao L, Amorim A, Carracedo A, Salas A. 2005. The genetic legacy of western Bantu migrations. *Hum Genet.* 117:366–375.
- Berniell-Lee G, Sandoval K, Mendizabal I, Bosch E, Comas D. 2007. SNPlexing the human Y-chromosome: a single-assay system for major haplogroup screening. *Electrophoresis.* 28:3201–3206.
- Bosch E, Calafell F, Santos FR, Perez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J. 1999. Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am J Hum Genet.* 65:1623–1638.
- Bosch E, Lee AC, Calafell F, Arroyo E, Henneman P, de Knijff P, Jobling MA. 2002. High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci Int.* 125:42–51.
- Cavalli-Sforza LL, editor. 1986. *African Pygmies.* Orlando (FL).
- Coia V, Destro-Bisol G, Verginelli F, Battaglia C, Boschi I, Cruciani F, Spedini G, Comas D, Calafell F. 2005. Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *Am J Phys Anthropol.* 128:678–681.
- Cruciani F, Santolamazza P, Shen P, et al. (15 co-authors). 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet.* 70:1197–1214.
- 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.
- Diamond J. 1997. *Guns, germs, and steel: the fates of human societies.* New York: W.W.Norton & Company.
- Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science.* 300:597–603.
- Goldstein DB, Zhivotovsky LA, Nayar K, Linares AR, Cavalli-Sforza LL, Feldman MW. 1996. Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. *Mol Biol Evol.* 13:1213–1218.
- Gusmao L, Sanchez-Diz P, Calafell F, et al. (42 co-authors). 2005. Mutation rates at Y chromosome specific microsatellites. *Hum Mutat.* 26:520–528.
- Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME. 2008. Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history. *Am J Phys Anthropol.* 137:316–323.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.
- Knight A, Underhill P, Mortensen H, Zhivotovsky L. 2003. African Y chromosome and mtDNA divergence provide insight into the history of click languages. *Curr Biol.* 13:464–473.
- Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet.* 74:532–544.
- Newman J. 1995. *The peopling of Africa: a geographic interpretation.* New Haven (CT): Yale University Press.
- Oslisly R. 1995. The middle Ogooué valley, Gabon: cultural changes and palaeoclimatic implications of the last four-millennia. *Azania.* 39–40:324–331.
- Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS. 1998. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet.* 62:420–434.
- Pereira L, Gusmao L, Alves C, Amorim A, Prata MJ. 2002. Bantu and European Y-lineages in Sub-Saharan Africa. *Ann Hum Genet.* 66:369–378.
- 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.
- Perrois L. 2006. *Fang.* Milan, Italy: 5 Continents Editions.
- Phillipson DW. 1993. *African archaeology.* Cambridge (MA): Cambridge University Press.
- 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, Quach H, Harmant C, et al. (22 co-authors). 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci USA.* 105:1596–1601.
- Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. *Am J Hum Genet.* 71:1082–1111.



- Salas A, Richards M, Lareu 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.
- Schneider S, Roessli D, Excoffier L. 2000. Arlequin ver. 2.0: a software for population genetic data analysis, 2.0 ed. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.
- Scozzari R, Cruciani F, Santolamazza P, et al. (16 co-authors). 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet.* 65:829–846.
- 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.
- 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.
- Underhill PA, Shen P, Lin AA, et al. (20 co-authors). 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet.* 26:358–361.
- Vansina J. 1995. New linguistic evidence and “the Bantu expansion”. *J African Hist.* 36:173–195.
- Verdu P, Austerlitz F, Estoup A, et al. (13 co-authors). 2009. Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Curr Biol.* 19:312–318.
- Wood ET, Stover DA, Ehret C, et al. (10 co-authors). 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet.* 13:867–876.
- Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339–348.

Lisa Matisoo-Smith, Associate Editor

Accepted March 30, 2009