BRILL Language Dynamics and Change 1 (2011) 50–88 brill.nl/ldc

Molecular Perspectives on the Bantu Expansion: A Synthesis*

Brigitte Pakendorf,[a] Koen Bostoen,[b] Cesare de Filippo[a]
a) Max Planck Research Group on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04 103 Leipzig, Germany
pakendorf@eva.mpg.de; cesare_filippo@eva.mpg.de

b) Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium
Ghent University, Department of African Languages and Cultures, Rozier 44, B-9000 Ghent, Belgium
Université libre de Bruxelles, Faculté de Philosophie et lettres, CP 175, avenue Fr. Roosevelt 50, 1050 Bruxelles, Belgium
koen.bostoen@africamuseum.be

Abstract

The expansion of Bantu-speaking peoples over large parts of sub-Saharan Africa is still a matter of debate—not only with respect to the propelling force behind it and the route(s) taken, but, also, in terms of the question whether there actually was a demographic expansion of peoples, rather than just a cultural expansion involving the spread of languages and technologies. In this paper, we provide a critical review of the extant linguistic and molecular anthropological data on Africa and discuss the insights they provide concerning the expansion itself as well as the demographic processes involved in it. Contrary to some assumptions by historians and cultural anthropologists, the genetic data speak in favor of an actual movement of peoples during the expansion of the Bantu languages over Africa, rather than a spread through language and culture shift. Furthermore, the molecular data indicate that sociocultural practices such as patrilocality and possibly even polygyny played a role in shaping the genetic diversity of Bantu-speaking peoples. These sociocultural practices might explain why, in Africa, there is a correlation between Y-chromosomal (i.e., paternal) lineages and linguistic affiliation, but not between mtDNA (maternal) lineages and language.

Keywords

Bantu languages; Bantu expansion; molecular anthropology; genetics

---

*) We gratefully acknowledge financial support by the Max Planck Society to B.P. and C. d. F, and funding by the Belgian Federal Science Policy to K.B. We thank Mark Stoneking for helpful discussion.

DOI: 10.1163/221058211X570349

1. Introduction

The dispersal of the Bantu language family is a topic that has incited a large amount of research, both amongst specialists in African history as well as from scholars working in other disciplines. Setting aside the higher phylogenetic levels of which they themselves are part, the Bantu languages constitute by far Africa's largest language family, both in terms of number of speakers and languages as well as in terms of geographical spread. About one African in three speaks a Bantu language, and estimates of the total number of languages vary between 440 (Guthrie, 1971) and ca. 680 (Mann and Dalby, 1987), mainly depending on where one draws the line between a language and a dialect (Nurse and Philippson, 2003a). As regards geographical distribution, the overall majority of communities living south of a line from southern Nigeria in the west to southern Somalia in the east and as far as the Cape are Bantu-speaking (Nurse and Philippson, 2003a). In spite of this numerical and geographical superiority, the Bantu family is widely recognized as forming only a low-level subgroup of the Niger-Congo phylum, being at most a subbranch of a branch of the Benue-Congo node (Schadeberg, 2003; Williamson and Blench, 2000). Blench (2006: 126) tentatively links the expansion of the Niger-Congo phylum with the improving climate at the beginning of the Holocene, i.e. from 12,000–10,000 BP onward. The beginning of the Bantu dispersal is usually situated much later. Vansina (1995: 52) proposes 5,000 BP on the basis of glottochronology. Contrasting the relatively low linguistic diversity within Bantu with the much higher diversity within Niger-Congo, Blench (2006: 134, 136) prefers a later date, 4,000 BP. As discussed in Bostoen (2007), this approximate time depth of 5,000–4,000 BP correlates with the definitive breakthrough of Neolithic technologies, such as macrolithic tools, polishing, and pottery, in the archaeological record of the Grassfields region of Cameroon (Lavachery, 2001). Slightly younger, but closely related technological assemblages have been recovered from several sites further south, i.e. in Cameroon, Gabon, the two Congos and the Central African Republic. These date from the early second millennium BCE until the last centuries BCE, and the younger they are, the further south(east) they are generally located. This gradual Neolithic expansion is commonly associated with the Bantu language dispersal (de Maret, 1994–1995). The striking contrast between this relatively shallow time depth and the current widespread distribution of the language family raises the key question of the 'Bantu Problem' (Eggert, 2005), namely, which factors were involved in the relatively rapid spread of this group of closely related languages over a disproportionately large part of sub-Saharan Africa (Bostoen, 2007; Eggert, 2005; Ehret, 2001; Vansina, 1979; 1995).

Historical linguists, archaeologists, and historians have had a long-standing debate in search of answers to this question. While linguists made the classifications available from which possible migration patterns could be derived, archaeologists provided insights into the cultural processes that may have enabled the Bantu expansion. The prevailing synthesis emerging from this protracted multidisciplinary quest for the driving forces behind the Bantu expansion is a model which identifies the dispersal of Bantu languages and the spread of the Early Iron Age industrial complex, consisting of pottery, metallurgy, domesticated plants and animals as well as a sedentary way of life, in sub-equatorial Africa as results of the same historical event—the migration of Bantu-speaking farmers (Huffman, 1980; Huffman and Herbert, 1994–1995; Oliver, 1966; Phillipson, 1977; 1985; 1989; Vansina, 1995). This model of the Bantu expansion has been proposed in different versions and the contents of the cultural package that accompanied the language spread vary from one proposal to another. In most models, metallurgy or grain cultivation are still absent during the first expansions through the equatorial rainforest, and only join the package when the dispersal in sub-equatorial Africa starts (de Maret, 1994–1995; Ehret, 1974; Vansina, 1990; 2006). This package model has been severely criticized for its lack of critical evaluation of underlying concepts and methods and for its propensity to circular argumentation (Eggert, 2005; Möhlig, 1989). Another recurrent criticism of the package model of the Bantu expansion concerns the overemphasis on migration as the single historical process underlying the spread of the Bantu languages (Gramly, 1978; Lwanga-Lunyiigo, 1976; Robertson and Bradley, 2000; Schepartz, 1988; Vansina, 1995). Nonetheless, this remains the reference model, especially for non-specialists, not least thanks to its explanatory elegance and straightforwardness.

Over the past 20 years, the debate has been joined by molecular anthropologists supplying an entirely new body of evidence that can be applied to the Bantu question and similar problems of language dispersal. As we discuss in this paper, recent developments in the field of molecular genetics have opened pathways that have the potential of leading to significant new insights into the demographic processes underlying the Bantu expansion. At the same time, unfortunately, many geneticists have been industrious in adding momentum to historical scenarios flawed by circular reasoning, rather than challenging earlier theories. Where Bantu speakers are involved, the Bantu migration model is often relied on for the historical interpretation of the geographic distribution of genetic markers. It is obvious that the risk of lapsing into assumptions that simply reinforce what has been stated before, instead of testing the original hypotheses, is particularly high: genetic data serve to reconstruct the demographic dynamics that may have underlain the Bantu language dispersal, yet it is precisely this preconceived Bantu migration model that is relied on to interpret the genetic data. Especially among proponents of the farming/language dispersal hypothesis, the Bantu expansion is seen as a textbook case of the concurrent dispersal of early agriculture, human genes, and languages (Bellwood and Renfrew, 2002; Renfrew, 1992; 2002). Diamond and Bellwood (2003: 598) express this vision in the plainest way possible when describing the Bantu language dispersal as one of the world's clearest examples of "expanding farmers bearing their own archaeologically visible culture, domesticates, skeletal types, genes, and languages."

Such bold claims are definitely not a reflection of the present state of knowledge. As regards the beginning and spread of agriculture in the Bantu area, Neumann (2005) convincingly demonstrates that independent material evidence from the archaeological record is very scarce, since unambiguous supporting data on crop remains or agricultural tools are largely missing. Concerning skeletal types, the Bantu expansion has been associated with the spread of the so-called 'Negro race' over central and southern Africa (Hiernaux, 1968; 1974). Such associations have justly been discarded, however, not only because it has become abundantly evident that there is no biological basis for distinguishing races amongst humans on the whole (Cartmill, 1998; Serre and Pääbo, 2004; Templeton, 1998), but also because, in the absence of written documents, archaeological remains do not allow any insights into the language spoken by the people whose remains are studied. Furthermore, there are no chronological or geographical grounds to link the distribution of alleged 'Negro' skeletal material with a particular language family (Lwanga-Lunyiigo, 1976; Robertson and Bradley, 2000), nor do mtDNA data "support the clustering of sub-Saharan Africans into (pre-Holocene) geographical races" (Salas et al., 2002: 1106).

A frequent problem that besets molecular anthropological studies is the implicit assumption that the modern-day distribution of ethnolinguistic groups is identical to their prehistoric distribution. However, it should be clear that the modern-day Bantu languages, which are presumed to have originated from one shared ancestor about 4,000 to 5,000 years ago as mentioned above, must have undergone a large amount of internal and contact-induced change to produce the striking diversity extant today. Thus, no single modern-day Bantu language can be equated with any ancestral language. Furthermore, ethnolinguistic communities are often fluid entities, shifting their allegiance over time. In Africa, most individuals are multilingual and so there is no one-to-one correlation between linguistic and genetic affiliation, which poses a problem for the investigation of language dispersals using current affiliations. These problems might appear to make the attempt at elucidating the past a fruitless waste of time and effort. However, the fact still remains that the very closely related Bantu languages are spread over a wide area of sub-Saharan Africa, and this geographic spread needs to be explained. While languages, culture and ethnic affiliation can change relatively rapidly, genetic turnover can only occur through immigration and replacement of the pre-existing populations, such as occurred in historical times in Tasmania. This relative temporal stability of genetic material enables molecular anthropologists to draw inferences about prehistoric demographic events, and these inferences can be integrated with archaeological and linguistic data to provide a more complete perspective on African history. Of course, as pointed out by MacEachern (2000), such inferences can only be as good as the data they are based on; flawed sample sets will lead to erroneous conclusions. Nevertheless, as shown by recent, carefully conducted studies (Quintana-Murci et al., 2008; Veeramah et al., 2010), the field has come a long way towards solving such issues.

In this interdisciplinary review paper we aim to provide a critical assessment of the available molecular genetic evidence for the Bantu migration and correlated demographic processes of interest to historical linguists. Since the Bantu language family is first of all a linguistic unit, we start by introducing the linguistic evidence for the spread of these languages.

2. Historical Classification(s) of the Bantu Languages and Hypotheses of Migration

The historical unity of the Bantu languages is widely recognized. Since Greenberg (1972), the region between southeastern Nigeria and western Cameroon has generally been accepted as the Bantu homeland. It is not only the zone of highest linguistic diversity within Bantu, but Bantu languages also meet there with several other subgroups of the larger Benue-Congo family; amongst others, their closest relatives known as Bantoid languages. For lack of indisputable innovations that delineate Bantu from these closest relatives, the demarcation of the Bantu area is problematic. It is impossible to assume a clear-cut split between 'Narrow' and 'Wide' Bantu languages, i.e., the Bantu languages as conventionally classified by Guthrie (1948; 1971) vs. the closely related 'Bantoid' languages from NW-Cameroon and SE-Nigeria. The two groups form a linguistic continuum rather than two clearly distinct branches of the Southern Bantoid subgroup of Benue-Congo, which is itself one of the main branches of the Niger-Congo phylum (cf. Blench, 2006: 120; Nurse and Philippson, 2003b).

As regards the internal classification of Bantu languages, most attempts have relied on lexicostatistics (Bastin, 1983; Bastin et al., 1979; 1999; Coupez et al., 1975; Heine, 1973; Heine et al., 1977; Henrici, 1973; Nurse and Philippson, 2003b). This method produces language trees based on the calculation of lexical distance between language pairs in terms of percentages of shared

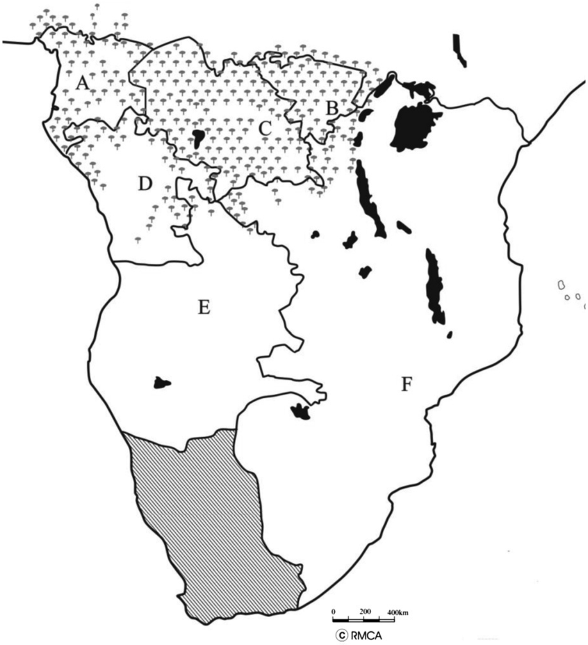

Figure 1. Approximate location of the major Bantu subgroups discussed in the text: A North-West, B Lebonya-Boan, C Inner Congo Basin, D West-Coastal, E South-West, F East.

cognates (cf. Nurse, 1997). Bantu classifications based on the classical comparative method, that is, on phonological or morphological innovatory features, are far less common and certainly do not result in the straightforward genealogical trees generated by lexicostatistics, due to the convergence effects to which Bantu languages have been exposed (Bastin, 1980; 1983; Ehret, 1999; Hyman, 1999; Möhlig, 1977; 1981; Nurse and Philippson, 2003b). Multilingualism as well as intensive and long-term contacts between Bantu speech communities have resulted in the transfer of features between closely related languages, which is often difficult to detect if it happened a long time ago. These convergence phenomena tend to blur linguistic frontiers and to complicate the demarcation of clear-cut Bantu subgroups. In addition, the internal relationships that exist between the subgroups and the way Bantu languages spread over their current territory are also a matter of debate. In contrast to what might be assumed, it is important to realize that no global proposal of internal Bantu language classification has yet received unanimous approval (Schadeberg, 2003).

It is generally acknowledged that there is considerably more historical diversity in the western part of the Bantu domain than in the eastern part. Bantu languages in the northwest, mostly in Cameroon and to a lesser extent in Gabon, are in many respects distinct from the other Narrow Bantu languages. Not only do they consistently turn up as a separate clade in lexicostatistical studies, they are also quite different typologically, e.g., in terms of maximality constraints on the verb stem (Hyman, 2004). For these reasons and because of the geographical proximity to the assumed Bantu homeland, they are generally considered as descendants of early breakaways from the Bantu nucleus (Bastin and Piron, 1999; Nurse and Philippson, 2003b; Vansina, 1995). Apart from a couple of isolated languages in the northeastern part of the DRC (the so-called Lebonya/Boan Bantu languages), the other western Bantu subgroups, i.e. Inner Congo Basin, West-Coastal Bantu and Southwest Bantu, are considered more closely related amongst each other than with the rest of Bantu languages and possibly go back to an intermediate common ancestor, sometimes labeled 'Narrow West Bantu' (Vansina, 1995). East Africa contains only the East Bantu subgroup, which is widely accepted as forming a coherent unit consisting of a large northeastern Savanna group in the north and several smaller groups in the south (Nurse and Philippson, 2003b: 175; see Fig. 1).

A major debate in historical Bantu studies concerns the question whether East Bantu is a primary Bantu branch or rather a later offshoot from a western Bantu node. This question is of great importance for our understanding of the Bantu language dispersal, seeing that two models of evolution prevail at present (Bostoen, 2004; Bostoen and Grégoire, 2007; Wiesmüller, 1997). According to the 'East separate from West' model, East Bantu is a primary branch of the Proto-Bantu node and its emergence in eastern Africa is the result of an eastward dispersion from Cameroon along the borders of the rainforest to the Great Lakes region. The spread of western Bantu is an entirely distinct event, characterized by a north-south movement through the equatorial rainforest and a rapid disintegration into different subgroups (Bastin et al., 1999; Coupez et al., 1975; Möhlig, 1981). In contrast, according to the 'East out of West' model, East Bantu is only a relatively late Bantu offshoot and constitutes a greater subgroup called 'Savanna Bantu' together with, roughly, the West-Coastal Bantu and Southwest Bantu subgroups, as opposed to the remainder of the Bantu languages, the so-called 'Forest Bantu' languages (Ehret, 2001; Heine et al., 1977; Henrici, 1973; Holden et al., 2005; Rexová et al., 2006; see Fig. 2). Roughly speaking, this latter scenario assumes an initial spread from the Bantu homeland through the tropical forest into the lower Congo region, from where a second wave would have started in different directions. One of these secondary migrations would be the origin of East Bantu. Several scholars, however, have argued against the historical validity of the so-called 'Forest' and 'Savanna' subgroups. Grégoire (2003) demonstrates that the forest is an area of great linguistic diversity, and Bantu languages spoken there do not

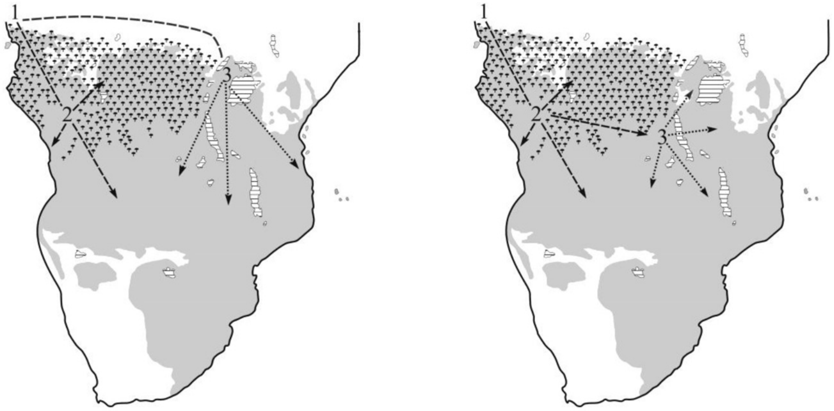

Figure 2. The two prevailing models of the Bantu expansion. 1 = Proto-Bantu nucleus, 2 = West-Bantu nucleus. 3 = East-Bantu nucleus. A: The 'East separate from West' model; B: The 'East out of West' model.

constitute a homogeneous subgroup that can be opposed to Savanna Bantu. Nurse and Philippson (2003b) have shown that several 'Savanna' languages from the western part of the Bantu range share several innovations with the 'Forest' languages to their north, which they do not have in common with the other 'Savanna' languages from the East. The current closeness between certain Savanna Bantu languages originally belonging to distinct eastern and western subgroups is most likely the result of sustained language contact involving profound convergence effects (see, for instance, Holden and Gray, 2006). As will be discussed in the following (cf. section 4.2), these contacts have also leveled the genetic differentiation between Bantu-speaking peoples in western and eastern Africa.

3. Molecular Anthropological Studies in Africa[1]

Like other biological anthropologists, molecular anthropologists study the differences (polymorphisms) between humans, focusing, however, not on the phenotypic level such as hair or skin color, but on differences at the molecular level. Under the assumption of a single origin of modern humans, similarities in the genetic make-up of individuals or populations may indicate

---

[1] For a basic introduction to molecular anthropology, readers are referred to Appendix 1 in Pakendorf (2007), as well as to the following specialist introductory literature: for the Y chromosome, Jobling and Tyler-Smith (2003); for mtDNA, Pakendorf and Stoneking (2005); as well as to the textbook by Jobling et al. (2004).

either shared ancestry or subsequent admixture, while differences can indicate that they have had separate histories. It should be noted that a 'population' in molecular anthropological studies is not a given, hard-and-fast concept. While a population in theoretical population genetics is a unit of randomly mating individuals, this does not necessarily hold for molecular anthropological studies, where the units under investigation are often defined by the research questions. In general, attempts are made at defining the populations of interest based on linguistic and geographic information as well as self-identification. However, the scale may differ widely, depending on the interest of the investigators (cf. MacEachern, 2000). Thus, a study dealing with fine-scaled differences between subgroups within one or two ethnolinguistic groups (e.g., Pakendorf et al., 2007) may analyze the genetic variation in units corresponding to one or two settlements, while investigations of broad-scale geographic differentiation (e.g., Lao et al., 2008; Novembre et al., 2008) may analyze populations at the state level. These divergent levels of analysis do not in themselves constitute a problem as long as it is kept in mind what the units of investigation are, and as long as similar units are compared with each other. Problems can and do arise, however, when data from smaller ethnolinguistic communities are taken to represent larger geographic or linguistic groups, without the inherent limitations being kept in mind. Thus, Tishkoff et al. (2009) use two small samples of 'South African Khoisan' (six Ju speakers and eight !Xun/Khoe, i.e. a mixed group) to represent the 'South African Khoisan' groups as a whole; rather than being aware of the fact that these small samples cannot possibly be representative of all Khoisan-speaking groups from southern Africa, they make questionable claims about African prehistory on this basis.

Overall, the African continent is still severely understudied from a molecular anthropological perspective. The recent increase in molecular anthropological studies in Africa notwithstanding, large areas, especially in central and southern Africa, still remain blank, and among the studies that do exist, very few have undertaken to analyze both mitochondrial DNA (mtDNA) and Y-chromosomal variation in the same population groups,[2] thus leading to somewhat one-sided conclusions. MtDNA has two major advantages for studies of population history: first of all, it is present in the cell in high numbers of largely identical molecules that do not undergo recombination. Therefore, mutations remain within the physical sequence of nucleotides in which they first arose, and later mutations occur on the background of earlier ones; this permits the reconstruction of evolutionary trees of mutations. Secondly, mtDNA is inherited solely in the maternal line, thus shedding light on the maternal (i.e. female)

---

[2] Some exceptions are: Beleza et al. (2005); Destro-Bisol et al. (2004b); de Filippo et al. (2010); Knight et al. (2003); Tishkoff et al. (2007); see also the Appendix.

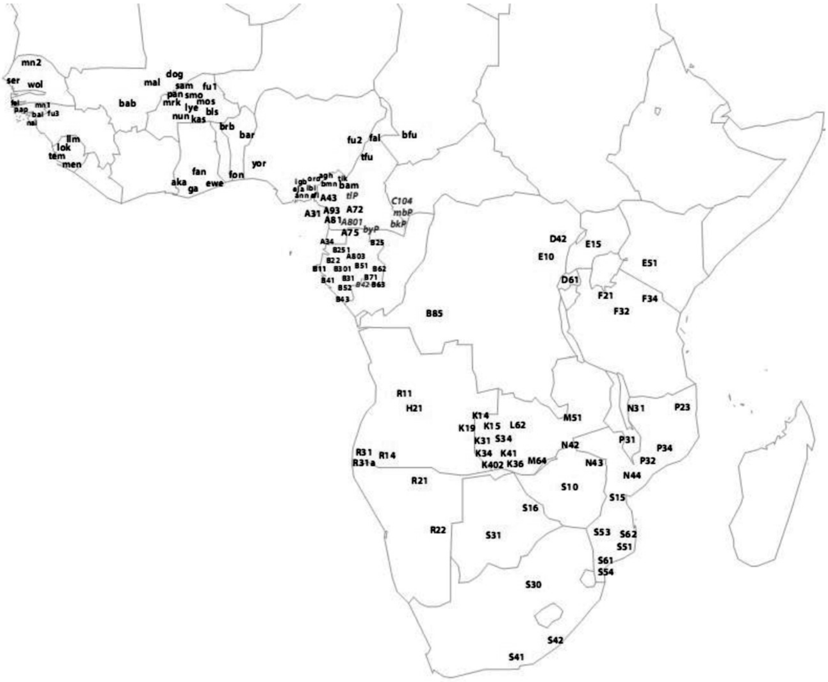

Figure 3. Map showing the approximate location of Niger-Congo samples included to date in molecular anthropological studies.

prehistory of the population. The counterpart of mtDNA is the Y chromosome, which has been the focus of numerous molecular anthropological studies in recent years. This chromosome, which determines the male sex of its carrier, is present only in men and is thus inherited only in the paternal line. Therefore, it illuminates the paternal history of a population. Like mtDNA, most of the Y chromosome does not undergo recombination,[3] since there is only one such chromosome present in each cell. This permits the reconstruction of evolutionary trees for mutations arising on the Y chromosome, and therefore mutations can be traced back in time.

A relatively large number of populations speaking languages classified as Narrow Bantu have been included in diverse studies of African genetic variation (cf. the Appendix, incl. references therein); however, geographic coverage is very patchy, with eastern Africa, southwestern Africa, and central Africa particularly poorly covered (see Fig. 3).

---

[3] In contrast to the mtDNA, however, which does not undergo recombination at all, a small part of the Y chromosome does recombine with the X chromosome.

Thanks to two recent studies of mtDNA and Y-chromosomal variation on the same individuals from Gabon and Cameroon (Quintana-Murci et al., 2008; Berniell-Lee et al., 2009), a large number of Northwest Bantu and West Coastal Bantu groups have now been analyzed. Eastern Africa is severely underrepresented with respect to Bantu-speaking groups, notwithstanding the importance of this region for investigating the hypotheses on the direction of Bantu migration. Only eight different Bantu-speaking ethnolinguistic groups have been included in molecular anthropological studies with sample sizes of 15 or more individuals, mostly with a focus on Y-chromosomal variation. Finally, southern Africa is by no means better represented, with the exception of Mozambique, from where several populations have been included in studies (Pereira et al., 2001;[4] 2002; Salas et al., 2002; Sikora et al., 2011). For south African Bantu-speaking groups, mtDNA data are practically nonexistent, with a couple of exceptions. One study, which focused on a specific deletion on the mtDNA, sequenced the so-called hypervariable region in only a subset of samples (Soodyall et al., 1996), while complete mitochondrial genome sequences are available for a handful of individuals in a worldwide comparison (Kivisild et al., 2006). On the other hand, several groups from this region, such as Zulu, Xhosa, Sotho, and Tswana, have been analyzed for Y-chromosomal markers (Wood et al., 2005).

In recent years, large-scale studies of the variation present in the so-called autosomes (non-sex chromosomes, i.e. all the chromosomes excluding the X and Y chromosomes) have grown in importance. The autosomes have the advantage that they represent the bulk of the human genome, thereby providing a large amount of polymorphisms that can be analyzed. However, the autosomes are inherited from both parents and undergo recombination; it is therefore impossible to trace individual lineages back in time. Thanks to the development of innovative computational methods, however, such large-scale autosomal data can now be used for the investigation of demographic events, such as changes in population size, as well as proportions of shared ancestry—be it due to inheritance or to admixture between groups. The most important study of African genetic diversity to date was conducted by Tishkoff et al. (2009), who studied over 1,000 autosomal markers in over 2,000 African samples from 121 populations speaking languages belonging to various families (and including some 'mixed populations,' such as the !Xun/Khoe, mentioned above, or the Tutsi/Hutu). However, here, too, coverage is patchy, with a relatively small number of Bantu samples coming predominantly from Cameroon, Tanzania and Kenya, and a dearth of data from central and southern Africa.

---

[4] This study, however, does not provide ethnolinguistic sample information.

This summary of molecular anthropological analyses of Bantu-speaking populations available to date has shown that there still exist large lacunae (cf. Fig. 3), especially with respect to populations from east, central, and southern Africa. Furthermore, the lack of data from both the Y chromosome and mtDNA makes it very difficult to come to comprehensive conclusions concerning the demographic processes involved in the Bantu expansion. It is therefore rather surprising that several papers refer to the Bantu expansion in their discussion or conclusions sections, generally claiming that the data presented support the assumption of such an event. As pointed out above, very few papers have undertaken a detailed study of the question, and in general the claims about the expansion are based on very scanty data and should be regarded with due caution. Nevertheless, as will be discussed in the following, even with the restricted amount of data currently available, it is possible to draw some conclusions regarding the relationships and history of Bantu-speaking populations.

4. Molecular Anthropological Perspectives on the Bantu Expansion

4.1. *Correlations between Linguistic and Genetic Relationships*

If one assumes a correlation between 'genes' and languages, then the classification of languages should reflect the genetic relationship of the groups speaking these languages to a certain extent. Such a correlation might be expected because, in general, children grow up learning the language of their parents, so that the transmission of genes and languages follows the same path. However, there are several factors that can confound such a correlation. Since humans can acquire any language, children may grow up speaking not the language of their parents, but that of the majority group surrounding them. Furthermore, human groups do not live in a vacuum, but come into contact with each other, which can have both genetic and linguistic effects. Population contact can lead to the interchange of genetic material through genetic admixture (intermarriage), confounding the regularity of correlations between the genetic and linguistic relationships of the groups concerned. Likewise, contact can lead to the interchange of linguistic material, be it copies of lexical or morphological items or copies of structural patterns. The most severe disruption of the correlation between genes and languages takes place when an entire ethnolinguistic group gives up its ancestral language and shifts to an unrelated language. However, although the effects of contact (which have certainly been at play in the Bantu languages, cf. section 2) can disrupt the correlation between the genetic and linguistic affiliation of a group on an individual scale, this connection might still hold at a broader level of comparison.

The correlation between linguistic and genetic affiliation in Africa was addressed by Scozzari et al. (1999) on the basis of Y-chromosomal data, Salas et al. (2002) on the basis of mtDNA data, and Wood et al. (2005) with a comparison of Y-chromosomal and mtDNA data. Scozzari et al. (1999) performed a very basic analysis of the amount of genetic variation found within populations grouped together by affiliation to one of the four Greenbergian language phyla (Afro-Asiatic, Niger-Congo, Nilo-Saharan, and Khoisan) as compared to the genetic variation found in large geographical groupings. They come to the conclusion that, overall, geography plays a larger role in structuring Y-chromosomal variation in Africa than linguistic affiliation does. A similar lack of correlation between genetic variation and linguistic affiliation was also found by Salas et al. (2002), who conclude that geography plays a larger role than linguistic affiliation in shaping mitochondrial DNA diversity in Africa. These results can be interpreted as evidence that marriage partners are chosen from geographically neighboring groups, irrespective of their linguistic affiliation—not entirely surprising, given the amount of multilingualism in Africa. In contrast, a very similar analysis by Wood et al. (2005), comparing Y-chromosomal with mtDNA data, leads to the result that while mtDNA variation does not correlate with linguistic affiliation, Y-chromosomal variation is better explained by linguistic affiliation to the four Greenbergian phyla than by geographical grouping. This finding is also supported by a finer-scale analysis performed on the basis of coarsely defined linguistic distances, suggesting "that African languages tend to be passed from father to children" (Wood et al., 2005: 873). In a recent study (de Filippo et al., 2011), we find a similar correlation between linguistic affiliation to the four major phyla and Y-chromosomal distances; however, the effects of geographical proximity (i.e., contact) are clearly discernible as well. This discrepancy between the correlation of linguistic affiliation with Y-chromosomal and mtDNA variation, respectively, can also be interpreted as evidence for the overall patrilocal social structure of African populations, in which brides are expected to move to their husbands' place of residence. In such a case, even if the mothers of children come from a group speaking a different language, the children are surrounded by people speaking the same language as their fathers and grow up acquiring this language as their primary one.

Interestingly, while this correlation between linguistic affiliation and Y-chromosomal variation holds at the level of the four major linguistic phyla, it does not hold at a lower level. Thus, in a fine-scale study of genetic variation in ethnolinguistic groups speaking languages classified as Cross River, Southern Bantoid, and Igbo from the Cross River region in southeastern Nigeria, Veeramah et al. (2010) find that there is no correlation between either mtDNA or Y-chromosomal variation and linguistic affiliation; instead, there is a high level of intermarriage between the groups. These results demonstrate that, at a small geographic scale, linguistic differences do not necessarily constitute a barrier to intermarriage, in good accordance with the high levels of multilingualism prevalent in Africa. Finer-scale molecular anthropological studies in different regions of Africa are necessary to obtain a better perspective on the degree of intermarriage and corresponding processes of language contact that have been at play in this continent.

4.2. *Genetic Markers and Evidence for the Bantu Migration*

One important factor that can shape the patterns of genetic variation in current-day populations is the fate of the pre-existing variation in the ancestral population. When a population splits into two or more daughter populations, each descendant group will receive some of the variation found in the ancestor. Depending on the geographic and cultural circumstances involved, the daughter populations may each receive a slightly different subset of the initial variation, leading to differences between them. Furthermore, random factors will cause the disappearance of some of the mutations that may initially have been present in the daughter populations. For instance, individuals carrying these polymorphisms might not reproduce, resulting in their genetic lineage dying with them, or men might have only daughters, or women might have only sons, resulting in the termination of their Y-chromosomal and mtDNA lineages, respectively. Such random processes affecting the genetic variation in a population are known as genetic drift. Genetic drift has a stronger effect in a small population than in a large one; thus, demographic factors such as population size play an important role in shaping genetic variation as well. An extreme case of genetic drift is the so-called founder event, in which a small group of individuals leaves the ancestral population and migrates to a different location. This founding group will carry only a small subset of the variation present in the entire ancestral population, and this subset will be present at high frequency in the resulting new population.

Thus, a notable reduction in genetic diversity can indicate prehistoric migration of a small group of founders. Long-distance expansions are expected to lead to serial founder events, with the genetic variation decreasing the further one moves from the point of origin, as seen in the expansion of modern humans out of Africa (Ramachandran et al., 2005). Should the expansion of the Bantu languages have involved an actual migration of people speaking these languages, we might expect to find signatures of serial founder events when moving further from the putative homeland in the Grasslands region of Cameroon. A cultural diffusion of the languages and agricultural technology without concomitant movement of people, on the other hand, would not result in such a reduction of genetic diversity. It is therefore noteworthy that modern-day Bantu-speaking groups are characterized by low Y-chromosomal diversity. Only two haplogroups are found in the majority of ethnolinguistic groups: E1b1a* and its subgroup E1b1a7.[5] A haplogroup is a group of related DNA molecules that is defined on the basis of shared mutations. This is similar to the linguistic innovations that are used by historical linguists to define nodes in trees depicting genealogical relationships between languages. The mutations defining specific haplogroups arise in a linear manner at different points in time on certain DNA molecules and spread through the population by reproduction; therefore, individuals carrying the same Y-chromosomal or mtDNA haplogroup are assumed to share a genetic ancestor. The distribution of particular haplogroups amongst communities speaking certain languages may thus be indicative of close genetic ties between these communities. However, there is no a priori evidence that would allow researchers to distinguish between sharing of haplogroups due to shared ancestry vs. genetic admixture (intermarriage).

The Y-chromosomal haplogroups E1b1a* and E1b1a7 account for 50–100 % of the Y-chromosomal variation in Bantu-speaking populations studied to date (Wood et al., 2005). E1b1a* is found in approximately 43 % of Bantu-speaking men, with a range from 17 % in the Hema to 66 % in the Ewondo. E1b1a7, on the other hand, is found on average in 30 % of Bantu-speaking men, with a range from 11 % in the Hema to 67 % in the Nande (Beleza et al., 2005; Cruciani et al., 2002; Knight et al., 2003; Wood et al., 2005). Even though these markers appear to be 'Bantu-specific' and have frequently been associated with the Bantu expansion (e.g. Berniell-Lee et al., 2009; Cruciani et al., 2002; Thomas et al., 2000; Zhivotovsky et al., 2004), neither of them is restricted in its distribution to Bantu-speaking populations;[6] it should be noted at this point, however, that genetic markers need not be restricted to only one group in order to be informative about that group's prehistory. As a matter of fact, markers are very rarely restricted to individual populations: a group may be characterized by high frequencies of a marker also found in other populations, especially if there was a founder event in their prehistory.

E1b1a* is found in very high frequency in west Africa in groups speaking languages classified as belonging to the Mande, Atlantic, Dogon, Gur, Kwa and Benue-Congo branches of Niger-Congo, such as Mandinka (79–82 %), Wolof (68 %), Mossi (67 %), Ewe (73 %), and others (Cruciani et al., 2002; Rosa et al., 2007; Wood et al., 2005); it is also found in non-Niger-Congo-speaking

---

[5] As a matter of fact, de Filippo et al. (2011) show that the two major haplogroups found amongst Bantu speakers are actually subgroups of E1b1a* and E1b1a7, called E1b1a8 and E1b1a7a, respectively.

[6] Cf. Appendix in Wood et al. (2005).

groups (Afro-Asiatic, Nilo-Saharan, and Khoisan). Although in some cases its presence in these communities might be due to admixture from Niger-Congo groups, the prevalence in Africa overall is so widespread that one cannot classify it as a Bantu-specific marker, nor as a marker characteristic of Niger-Congo speaking groups (cf. Quintana-Murci et al., 2010).

E1b1a7, on the other hand, which is a sub-lineage of E1b1a*, although not restricted to Bantu-speaking groups, does have a fairly restricted distribution in Africa. It is found predominantly in communities speaking languages of the Kwa and Benue-Congo branches of the Niger-Congo phylum, with the few exceptions being groups that may have obtained it through admixture with Bantu-speaking groups (e.g. the Mbuti pygmies or the Sekele !Kung; Cruciani et al., 2002; Wood et al., 2005). Thus, the distribution of this haplogroup in African populations provides some indication that the expansion of the Bantu-speaking groups over sub-Saharan Africa may have involved actual migration, not just the diffusion of languages and technology.

However, the genetic diversity associated with the sub-haplogroups E1b1a8 and E1b1a7a mentioned in footnote 5 does not diminish with distance from the putative homeland. This can be established with analyses of a different type of polymorphism, called short tandem repeats (STRs, also known as microsatellites). As indicated by the name, STRs are stretches of short repeated sequences between two and six base pairs long. DNA repetitions like these pose problems for the reproductive apparatus of the cell, leading to a high error rate in replication. Because of the high error (= mutation) rate of STRs, these are in general highly variable between individuals, permitting the investigation of the diversity present within haplogroups. Because the STR diversity within sub-haplogroups E1b1a8 and E1b1a7a does not diminish with geographic distance from the Bantu homeland, there is no evidence of the serial founder event one might expect if groups of people moved progressively through sub-Saharan Africa (de Filippo et al., 2011). This lack of evidence for migration might be interpreted as supporting the hypothesis of cultural diffusion, rather than actual migration of peoples, after all. What contradicts this suggestion is that there are no differences between Bantu-speaking groups in eastern Africa and those in western and central Africa with respect to the STR variation within these sub-haplogroups. This sharing of STR types between geographically distant Bantu groups would not be expected if they were originally indigenous groups who had only recently switched to the Bantu language and mode of subsistence. In such a case one would expect larger differences between western and eastern African Bantu-speaking groups, unless the shifting indigenous groups of western and eastern Africa were themselves closely related genetically—which, however, presupposes yet another recent spread of peoples over the continent, prior to the Bantu expansion. Thus, the

Y-chromosomal data do indicate actual migration of populations during the expansion of the Bantu languages. Combined with this evidence for actual movement of peoples, the lack of serial founder events indicates that either the migrating groups were large enough that there was no reduction in genetic diversity, or that the signal of an original founder event has subsequently been erased by later migrations—in good accordance with the linguistic evidence for contact (cf. section 2).

MtDNA diversity, on the other hand, is quite high in Bantu-speaking populations, even in those groups settled at the further end of the putative migration, such as Mozambique or Zambia (de Filippo et al., 2010; Salas et al., 2002). This high diversity can be interpreted as an indication that there were successive waves of genetically diverse migrations, or that the Bantu expansion involved a large group of migrants that took their origin in a diverse ancestral population. An alternative explanation for the relatively high mtDNA diversity in Bantu-speaking groups is that intermarriage with mostly women from indigenous communities took place in the course of the Bantu expansion, as will be discussed in more detail in section 4.3.2 below—that is, that the expansion involved predominantly migration of men.

Nevertheless, there is some indication that the dispersal of Bantu-speaking groups to southeastern Africa involved a founder effect. This is seen in the high frequency and reduced diversity of both mtDNA haplogroups L0a1a and L0a2 in the Mozambican Bantu-speaking groups (Salas et al., 2002: their Fig. 4) as well as haplogroups L2a1a and L2a1b (Salas et al., 2002: their Fig. 6). L2a1b in particular is interesting because it shows signs of expansion after the founder effect: a high-frequency haplotype from which a number of haplotypes that differ by only one mutation are derived (Salas et al., 2002: their Fig. 6). This is corroborated by the fact that the sequence diversity in Mozambique (0.96 ± 0.01) is slightly lower than that found in most west or east African populations. More investigations as detailed as that undertaken by Salas et al. (2002) are needed for other regions of Africa to elucidate the demographic factors at play during the course of the Bantu expansion.

In their study of autosomal variation, Tishkoff et al. (2009) find a strong signal of genetic relationship of the Niger-Congo groups throughout sub-Saharan Africa, possibly reflecting the spread of Bantu-speaking peoples. This signal is due to similar allele frequencies of the polymorphic markers investigated and not to the persistence of a specific lineage, such as the Y-chromosomal and mtDNA haplogroups discussed above. In contrast to the lack of differentiation within the Y-chromosomal sub-haplogroups E1b1a8 and E1b1a7a between western and eastern African Bantu-speaking groups, as discussed above, they also find weaker signals for "subtle substructure between East African

Bantu-speakers […] and West-Central African Bantu-speakers […], and individuals from Nigeria and further west, who speak various non-Bantu Niger-Kordofanian languages" (Tishkoff et al., 2009: 1039). Interestingly, in a recent study using ~2,800 autosomal markers in a set of 11 populations of African descent, Sikora et al. (2011) come to the conclusion that the expansion of Bantu languages to southeastern Africa (Mozambique) may well have taken place not so much through actual migration of peoples as through the assimilation of local groups who shifted to the Bantu language. The discrepancy between the findings of Tishkoff et al. (2009) and Sikora et al. (2011) might be due to the different populations included—notably, the study by Tishkoff et al. lacks Mozambican populations entirely, and populations from southern Africa as a whole are severely underrepresented. On the other hand, the study by Sikora et al. (2011) is based on a very restricted sample of Bantu-speaking groups: other than 180 samples from Mozambique, which are not ethnolinguistically defined, they include published data on two very small and undefined samples of 'Bantu' from Kenya and South Africa, as well as a sample of Luhya. These samples are clearly not representative of the genetic variation present in Bantu-speaking populations. Another factor that may explain the divergent results is the differences in markers analyzed by the two studies: Tiskhoff et al. investigated a large number of highly variable STRs that play no functional role in the genome, while Sikora et al. studied slowly evolving sites in genes associated with immunity to diseases and inflammation response. The latter are expected to underlie selective pressures.

Although Salas et al. (2002) discuss their results in terms of the 'East separate from West' model (cf. section 2), while Plaza et al. (2004) suggest that the western and eastern streams of migration came together south of the rainforest and spread jointly throughout southern Africa, the preceding discussion should have made it clear that there is not yet sufficient data to warrant any conclusion concerning the potential route(s) of migration taken by the ancestors of modern Bantu-speaking groups. As described in section 3, there are still hardly any data from central, southwestern, eastern, or southern Africa. Thus, the genetic investigation of the putative routes of the Bantu expansion will have to be suspended until sufficient data have been collected. However, as mentioned above, the lack of genetic differentiation between Bantu-speaking groups from eastern Africa and central and western Africa with respect to the sub-haplogroups E1b1a8 and E1b1a7a may indicate that subsequent smaller migrations between Western and Eastern Bantu groups have erased the genetic signal of the initial split (de Filippo et al., 2011).

4.3. *Genetic Perspectives on Sociocultural Parameters Involved in the Bantu Dispersal*

4.3.1. Assimilation of Non-Bantu Speech Communities

A topic that has generated much interest among molecular anthropologists is the degree of admixture between the putative immigrating Bantu-speaking groups and the resident populations. Certain modern African hunter-gatherers are characterized by specific mtDNA and Y-chromosomal polymorphisms; therefore prehistoric admixture between incoming agriculturalists and resident hunter-gatherer groups genetically resembling modern foragers is detectable. For example, western pygmy groups are characterized by very high frequencies of mtDNA haplogroup L1c1a (~84%; Quintana-Murci et al., 2008), while southern African groups speaking Ju languages are characterized by haplogroups L0d and L0k. Similarly, the Y-chromosomal lineage A3b1 is characteristic of southern African Ju-speaking groups (Cruciani et al., 2002; Wood et al., 2005), while haplogroup B2b is more generally characteristic of African hunter-gatherer groups (Güldemann and Stoneking, 2008). However, as yet, relatively few hunter-gatherer groups have been included in molecular anthropological studies in Africa, so the prevalent idea of what would constitute 'hunter-gatherer introgression' is potentially biased and might lead to wrong conclusions.

From its distribution in current African populations, one could hypothesize that some mtDNA sub-clades of haplogroup L1c1a entered Bantu-speaking populations through admixture with central African pygmy groups: it is found in exceedingly high frequencies in western pygmies, as noted above, and is otherwise restricted to Bantu-speaking groups, with the exception of the Chadic-speaking Mandara (Coia et al., 2005) (5%) as well as the Yoruba, whose language is part of the Benue-Congo branch of Niger-Congo (6% of the sequences reported by Watson et al., 1996 and Vigilant et al., 1991). However, contrary to this hypothesis of admixture between the ancestors of the Bantu-speaking groups and the ancestors of modern-day pygmy groups, Quintana-Murci et al. (2008) propose that L1c was already present in the shared ancestor of agricultural groups (in their study represented by Bantu-speaking groups from Gabon and Cameroon) and pygmy groups; in addition, they suggest gene flow to have taken place over a long period of time after an extended period of initial isolation.

In southwestern Africa (Angola and the Angolan enclave Cabinda) there is no evidence for introgression of lineages characteristic of Ju-speaking groups (mitochondrial haplogroups L0d or L0k, Y-chromosomal A3b1 or B2b), which has been interpreted to indicate a replacement of local hunter-gatherers by immigrating Bantu speakers (Beleza et al., 2005; Plaza et al., 2004). However, as mentioned above, our knowledge of the genetic composition of African hunter-gatherer groups is as yet very limited and such conclusions cannot be regarded as definitive. For instance, if the local hunter-gatherer groups had a genetic composition very different from the Ju-speaking !Kung groups used as the prototypical 'Khoisan hunter-gatherers' in molecular anthropological studies, but similar to that of the immigrating groups,[7] admixture between the immigrating Bantu speakers and resident groups would not be detectable with the methods currently employed. Furthermore, since there is as yet no evidence that populations related to southern African Ju- or Tuu-speaking groups were formerly settled much further north than their current territory (Tom Güldemann, pers. comm.), lack of Ju-specific genetic markers in the immigrant Bantu communities does not necessarily provide evidence for a lack of admixture between the immigrants and local hunter-gatherers. On the other hand, the presence of clicks in some Bantu languages of southwestern Zambia demonstrates that some contact must have taken place in southwestern Africa (Bostoen and Sands, to appear).

In southeastern Africa, on the other hand, the Bantu-speaking groups appear to have intermarried to a minor degree with local hunter-gatherer groups that were genetically related to modern-day Ju-speakers, as evidenced by some level of 'hunter-gatherer specific' haplogroups (such as mtDNA L0d and Y-chromosomal A3b1) found in Bantu-speaking groups (de Filippo et al., 2010; Pereira et al., 2001; Salas et al., 2002; Wood et al., 2005). Interestingly, Southeast Bantu groups (Zulu, Xhosa, Sotho, and Tswana) have 28.5% of haplogroup L0d (Schlebusch et al., 2009); this is in good accordance with the linguistic facts, since the speakers of Zulu and Xhosa have adopted quite an extensive number of click phonemes into their languages (Güldemann and Stoneking, 2008: their Table 2). Thus, both the linguistic and the genetic data confirm that some of the Bantu-speaking groups who migrated to south Africa entered into close cultural and physical contact with the indigenous, click-using hunter-gatherers. However, while there is evidence of intermarriage with resident communities in southern Africa, the genetic data do not provide any indication that the spread of the Bantu languages and culture in this region was due only to shift of indigenous peoples. In such a case, we would expect to find a majority of forager-specific lineages in the Bantu-speaking groups, and not only a third or less. This, too, provides evidence that the Bantu expansion involved migration of people rather than just cultural diffusion, at least in this part of the continent.

---

[7] Cf. the difference between the 'Vasikela Kung' and the Khwe (Chen et al., 2000).

The contact between Bantu-speaking groups and hunter-gatherers is suggested to have been largely sex-biased, with a higher degree of female migration from hunter-gatherer groups into agricultural groups, and a higher degree of male migration from agricultural groups to the hunter-gatherers. This is accounted for by the fact that it is considered unacceptable for women from agriculturalist communities to marry men from foraging communities, while the reverse is more acceptable. Thus, hunter-gatherer mtDNA lineages can enter agriculturalist communities; conversely, if such unions break up and the woman returns to her parental community with her children, any sons by her agriculturalist husband will be carrying his Y chromosome, which then gets incorporated into the gene pool of the foraging community (Cavalli-Sforza, 1986). This sex-biased admixture pattern between farmers and foragers has been confirmed by several studies (Destro-Bisol et al., 2004b; Quintana-Murci et al., 2008; Tishkoff et al., 2007; Wood et al., 2005). The fact that such socio-cultural boundaries to intermarriage are not watertight, however, is shown by the south African Sotho, Tswana, Zulu, and Xhosa, which have a frequency of 3–7 % of the 'Ju-specific' Y-chromosomal haplogroup A3b1 (Wood et al., 2005). But since these groups have incorporated four to nine times as many hunter-gatherer female lineages (28.5 % on average), as described above, here, too, the gene flow from indigenous groups to immigrating Bantu speech communities was predominantly female-mediated.

4.3.2. Demographic Features of the Bantu Migrations

Genetic variation on the mtDNA and Y chromosome is strongly shaped by cultural factors such as social organization, polygyny and marriage practices (Ségurel et al., 2008). In most of the worldwide populations studied by molecular anthropologists up to now, the genetic diversity of the Y chromosome is smaller than that of mtDNA, and the genetic differences between populations are larger when Y-chromosomal markers are analyzed than when mtDNA is used. One reason for this could be the differential reproductive success of men in polygynous societies, where a few men can afford to have several wives and thus father more children than the majority of men, who can only afford one wife, if any, and thus father fewer or no children (Hammer et al., 2008; Kayser et al., 2003). Another reason could be that the vast majority of human societies have patrilocal residence patterns so that the women (and with them their mtDNA) move to their husbands' homes, while the men (and their Y chromosomes) stay in the village in which they were born (Oota et al., 2001; Seielstad et al., 1998). This leads to an incrementally higher amount of mixing of mtDNA than Y-chromosomal lineages, and results in higher mtDNA diversity.

Although it is difficult to compare values of mtDNA and Y-chromosomal diversity directly, since the underlying markers are ascertained in very different ways, there is still a general trend for mtDNA diversity amongst Bantu-speaking populations to be larger than the Y-chromosomal diversity (cf. section 4.2). As stated by Salas et al. (2002: 1105), most of the mtDNA haplogroups found in sub-Saharan Africa as a whole are found in the Bantu-speaking groups from Mozambique. On the other hand, all Bantu-speaking populations are characterized by very high frequencies of only two Y-chromosomal haplogroups, E1b1a* and its derivative E1b1a7[8] (Wood et al., 2005). Furthermore, genetic differences between populations are larger when using Y-chromosomal markers than mtDNA, suggesting that the paternal lineages are mixing less than the maternal lineages. These discrepancies between mtDNA and Y-chromosomal variation might be indicative of sex-biased sociocultural practices such as patrilocality, which may have affected the demographies of men and women in different ways (Hammer et al., 2008; Kayser et al., 2003; Oota et al., 2001; Pilkington et al., 2008; Seielstad et al., 1998). As discussed above, the Y-chromosomal evidence speaks in favor of actual migration of peoples in the course of the expansion of Bantu languages and agricultural technology. The differences between Y-chromosomal and mtDNA variation in modern-day Bantu-speaking populations imply that, when the ancestors of the Bantu speakers migrated, they took in wives from the local groups but not husbands, following patrilocal marriage rules. In this way, the distinctive Y-chromosomal composition of the migrating Bantu speakers stayed relatively stable, while new mtDNA lineages were continuously added to the Bantu gene pool. This is also shown by the incorporation of more forager mtDNA lineages than Y chromosomes in south African Bantu groups, as discussed above.

It has been suggested that the paternal gene pool of the Bantu-speaking groups may further have remained quite small through the practice of polygyny (cf. Wood et al., 2005). A man who has several wives can expect to father more offspring than a man who has only one or no wife; therefore, over time the genetic lineages of the few men with more than one wife will increase in frequency. This, conversely, leads to a decrease in genetic diversity of the Y chromosome, since the same Y-chromosomal lineage is being passed on to proportionally more sons, who in turn will pass it on to a large number of grandsons, great-grandsons, and so forth. While the differences in Y-chromosomal and mtDNA diversity are not nearly as strong in Africa as they are in the highlands of western New Guinea (Kayser et al., 2003), the Y-chromosomal diversity tends to be lower, providing some indication that polygynous practices might

---

[8] Or, as shown by de Filippo et al. (2011), of the sub-haplogroups E1b1a8 and E1b1a7a.

have played a role. However, this hypothesis remains to be tested by comparing Y-chromosomal diversity values between polygynous and non-polygynous societies.

5. Conclusions

The above discussion of the extant molecular anthropological data on African populations, especially those speaking Bantu languages, has shown that sweeping conclusions concerning the nature of the 'Bantu migration(s)' are premature, and that current efforts at elucidating the demographic processes and patterns of migration are hampered by a severe shortage of data as well as the lack of comparability amongst them.

However, the current state of molecular anthropological studies on African populations indicates that the Bantu language dispersal actually involved human migration, not just the diffusion of languages and technology. Y-chromosomal diversity in Bantu-speaking populations is lower than mtDNA diversity, with only two Y-chromosomal haplogroups present in high frequency among most Bantu speech communities: E1b1a8 and E1b1a7a. Although neither of them is a Bantu-specific marker, and they are not even restricted to Niger-Congo-speaking populations, their current distribution in African populations suggests that male migrants carrying these haplogroups were involved in the spread of Bantu languages.

On the other hand, the higher mtDNA diversity among Bantu-speaking populations is indicative of sex-biased sociocultural customs such as patrilocality, whereby Bantu-speaking groups commonly intermarried with local women during their expansion, rather than incorporating indigenous men. The presence of characteristic central African 'pygmy' mtDNA haplogroups, such as L1c1a, and mtDNA haplogroups characteristic of southern African Ju-speaking groups, such as L0d, in the maternal gene pool of certain Bantu-speaking groups provides evidence of such female-mediated gene flow from local communities into migrating Bantu speech communities.

To be sure, until coverage of central, eastern and southern Africa has been ameliorated, these conclusions can only be viewed as provisional. Nevertheless, molecular anthropological studies, and especially recent revolutionary innovations in the available methods, have the potential of providing novel insights into the demographic processes that underlie the expansion of the Bantu languages over much of sub-Saharan Africa. Thus, novel techniques that investigate variation at a vast number of polymorphisms in the entire genome can now be coupled with simulation-based analyses of the data, which may provide more definitive insights into the genetic history of Bantu populations.

References

Bastin, Yvonne. 1980. Statistique grammaticale et innovations en bantou. In Luc Bouquiaux (ed.), *L'expansion bantoue. Actes du Colloque International du CNRS, Viviers (France), 4–16 avril 1977*, vol. 2, 387–400. Paris: Société des Etudes Linguistiques et Anthropologiques de France.

Bastin, Yvonne. 1983. Essai de classification de quatre-vingts langues bantoues par la statistique grammaticale. *Africana Linguistica* 9: 11–108.

Bastin, Yvonne, André Coupez, and Bernard de Halleux. 1979. Statistique lexicale et grammaticale pour la classification des langues bantoues. *Bulletin des séances de l'Académie royale des sciences d'Outre-Mer* 3: 375–387.

Bastin, Yvonne, André Coupez, and Michael Mann. 1999. *Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study*. Tervuren: Royal Museum for Central Africa.

Bastin, Yvonne and Pascale Piron. 1999. Classifications lexicostatistiques: bantou, bantou et bantoïde. De l'intérêt des 'groupes flottants'. In Jean-Marie Hombert and Larry M. Hyman (eds.), *Bantu Historical Linguistics: Theoretical and Empirical Perspectives*, 149–164. Stanford: CSLI Publications.

Beleza, Sandra, Leonor Gusmao, Antonio Amorim, Angel Carracedo, and Antonio Salas. 2005. The genetic legacy of western Bantu migrations. *Human Genetics* 117: 366–375.

Bellwood, Peter and Colin Renfrew (eds.). 2002. *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research.

Berniell-Lee, Gemma, Francesc Calafell, Elena Bosch, Evelyne Heyer, Lucas Sica, Patrick Mouguiama-Daouda, Lolke van der Veen, Jean-Marie Hombert, Lluis Quintana-Murci, and David Comas. 2009. Genetic and demographic implications of the Bantu expansion: Insights from human paternal lineages. *Molecular Biology and Evolution* 26: 1581–1589.

Blench, Roger. 2006. *Archaeology, Language and the African Past*. Lanham: Altamira Press.

Bostoen, Koen. 2004. Linguistics for the use of African history and the comparative study of Bantu pottery vocabulary. *Antwerp Papers in Linguistics* 106: 131–154.

Bostoen, Koen. 2007. Pots, words and the Bantu problem: On lexical reconstruction and early African history. *Journal of African History* 48: 173–199.

Bostoen, Koen and Claire Grégoire. 2007. La question bantoue: bilan et perspectives. *Mémoires de la Société de Linguistique de Paris* 15: 73–91.

Bostoen, Koen and Bonny Sands. To appear. Clicks in south-western Bantu languages: contact-induced vs. language-internal lexical change. In Matthias Brenzinger (ed.), *Proceedings of the 6th World Congress of African Linguistics Cologne 2009*. Köln: Rüdiger Köppe Verlag.

Caglià, Alessandra, Sergio Tofanelli, Valentina Coia, Ilaria Boschi, Marina Pescarmona, Gabriella Spedini, Vincenzo Pascali, Georgio Paoli, and Giovanni Destro-Bisol. 2003. A study of Y-chromosome microsatellite variation in sub-Saharan Africa: A comparison between F(ST) and R(ST) genetic distances. *Human Biology* 75: 313–330.

Cartmill, Matt. 1998. The status of the race concept in physical anthropology. *American Anthropologist* 100: 651–660.

Castrì, Loredana, Sergio Tofanelli, Paolo Garagnani, Carla Bini, Xenia Fosella, Susi Pelotti, Giorgio Paoli, Davide Pettener, and Donata Luiselli. 2009. mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: Implications for peopling and migration patterns in sub-Saharan Africa. *American Journal of Physical Anthropology* 140: 302–311.

Cavalli-Sforza, Luca Luigi. 1986. *African Pygmies*. Orlando: Academic Press.

Černý, V., A. Salas, M. Hájek, M. Zaloudková, and R. Brdicka. 2007. A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: A history revealed by the mitochondrial DNA genome. *Annals of Human Genetics* 71: 433–452.

Chen, Yu-Sheng, Antonel Olckers, Theodore G. Schurr, Andreas M. Kogelnik, Kirsi Huoponen, and Douglas C. Wallace. 2000. mtDNA variation in the South African Kung and Khwe—and their genetic relationships to other African populations. *American Journal of Human Genetics* 66: 1362–1383.

Coelho, Margarida, Fernanda Sequeira, Donata Luiselli, Sandra Beleza, and Jorge Rocha. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evolutionary Biology* 9: 80.

Coia, Valentina, Alessandra Caglià, Barbara Arredi, Francesco Donati, Fabŕicio R. Santos, Arpita Pandya, Luca Taglioli, Giorgio Paoli, Vincenzo Pascali, Gabriella Spedini, Giovanni Destro-Biso, and Chris Tyler-Smith. 2004. Binary and microsatellite polymorphisms of the Y-chromosome in the Mbenzele pygmies from the Central African Republic. *American Journal of Human Biology* 16: 57–67.

Coia, Valentina, Giovanni Destro-Bisol, Fabio Verginelli, Cinzia Battaggia, Ilaria Boschi, Fulvio Cruciani, Gabriella Spedini, David Comas, and Francesc Calafell. 2005. Brief communication: mtDNA variation in North Cameroon: Lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *American Journal of Physical Anthropology* 128: 678–681.

Coupez, André, E. Evrard, and Jan Vansina. 1975. Classification d'un échantillon de langues bantoues d'après la lexicostatistique. *Africana Linguistica* 6: 131–158.

Cruciani, Fulvio, Piero Santolamazza, Peidong Shen, Vincent Macaulay, Pedro Moral, Antonel Olckers, David Modiano, Susan Holmes, Giovanni Destro-Bisol, Valentina Coia, Douglas C. Wallace, Peter J. Oefner, Antonio Torroni, L. Luca Cavalli-Sforza, Rosaria Scozzari, and Peter A. Underhill. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *American Journal of Human Genetics* 70: 1197–1214.

de Filippo, Cesare, Patricia Hein, Lawrence Barham, Mark Stoneking, and Brigitte Pakendorf. 2010. Genetic perspectives on the interactions between farmers and hunter-gatherers in the Luangwa Valley of Zambia. *American Journal of Physical Anthropology* 141: 382–394.

de Filippo, Cesare, Chiara Barbieri, Mark Whitten, Sununguko W. Mpoloka, Ellen D. Gunnarsdóttir, Koen Bostoen, Terry Nyambe, Klaus Beyer, Henning Schreiber, Peter de Knijff, Donata Luiselli, Mark Stoneking, and Brigitte Pakendorf. 2011. Y-chromosomal variation in Sub-Saharan Africa: Insights into the history of Niger-Congo groups. *Molecular Biology and Evolution* 28: 1255–1269.

de Maret, Pierre. 1994–1995. Pits, pots and the far west streams. *Azania* 29–30: 318–323.

Destro-Bisol, Giovanni, Valentina Coia, Ilaria Boschi, Fabio Verginelli, Alessandra Caglià, Vincenzo Pascali, Gabriella Spedini, and Francesc Calafell. 2004a. The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *American Naturalist* 163: 212–226.

Destro-Bisol, Giovanni, Francesco Donati, Valentina Coia, Ilaria Boschi, Fabio Verginelli, Alessandra Caglià, Sergio Tofanelli, Gabriella Spedini, and Cristian Capelli. 2004b. Variation of female and male lineages in sub-Saharan populations: The importance of sociocultural factors. *Molecular Biology and Evolution* 21: 1673–1682.

Diamond, Jared and Peter Bellwood. 2003. Farmers and their languages: The first expansions. *Science* 300: 597–603.

Eggert, Manfred. 2005. The Bantu problem and African archaeology. In Ann Brower Stahl (ed.), *African Archaeology: A Critical Introduction*, 301–326. Malden, MA: Blackwell.

Ehret, Christopher. 1974. Agricultural history in Central and Southern Africa, ca. 1000 B.C. to A.D. 500. *Transafrican Journal of History* 4: 1–25.

Ehret, Christopher. 1999. Subclassifying Bantu: The evidence of stem morpheme innovations. In Jean-Marie Hombert and Larry M. Hyman (eds.), *Bantu Historical Linguistics: Theoretical and Empirical Perspectives*, 43–147. Stanford: CSLI Publications.

Ehret, Christopher. 2001. Bantu expansions: Re-envisioning a central problem of early African history. *The International Journal of African Historical Studies* 34: 5–27.

Forster, Peter, Arne Röhl, Petra Lünnemann, Catrin Brinkmann, Tatiana Zerjal, Chris Tyler-Smith, and Bernd Brinkmann. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. *American Journal of Human Genetics* 67: 182–196.

González, A.M., V.M. Cabrera, J.M. Larruga, A. Tounkara, G. Noumsi, B.N. Thomas, and J.M. Moulds. 2006. Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations. *Annals of Human Genetics* 70: 631–657.

Gramly, Richard Michael. 1978. Expansion on Bantu-speakers versus development of Bantu language in situ. An archaeologist's perspective. *South African Archaeological Bulletin* 33: 107–112.

Graven, Laurent, Giuseppe Passarino, Ornella Semino, Pierre Boursot, Silvana Santachiara-Benerecetti, André Langaney, and Laurent Excoffier. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Molecular Biology and Evolution* 12: 334–345.

Greenberg, Joseph. 1972. Linguistic evidence regarding Bantu origins. *Journal of African Languages and Linguistics* 13: 189–216.

Grégoire, Claire. 2003. The Bantu languages of the forest. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 349–370. London: Routledge.

Güldemann, Tom and Mark Stoneking. 2008. A historical appraisal of clicks: A linguistic and genetic population perspective. *Annual Review of Anthropology* 37: 93–109.

Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. London: Oxford University Press for the International African Institute.

Guthrie, Malcolm. 1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages. Volume 2: Bantu Prehistory, Inventory and Indexes*. London: Gregg International.

Hammer, Michael F., Fernando L. Mendez, Murray P. Cox, August E. Woerner, and Jeffrey D. Wall. 2008. Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genetics* 4: e1000202.

Heine, Bernd. 1973. Zur genetischen Gliederung der Bantu-Sprachen. *Afrika und Übersee* 56: 164–185.

Heine, Bernd, Hans Hoff, and Rainer Vossen. 1977. Neuere Ergebnisse zur Territorialgeschichte der Bantu. In Wilhelm J.G. Möhlig, Franz Rottland, and Bernd Heine (eds.), *Zur Sprachgeschichte und Ethnohistorie in Afrika*, 52–72. Berlin: Dietrich Reimer.

Henrici, Alick. 1973. Numerical classification of Bantu languages. *African Language Studies* 14: 82–104.

Hiernaux, Jean. 1968. Bantu expansion: The evidence from physical anthropology confronted with linguistic and archaeological evidence. *Journal of African History* 9: 505–515.

Hiernaux, Jean. 1974. *The People of Africa*. London: Weidenfeld and Nicolson.

Holden, Claire J. and Russell D. Gray. 2006. Rapid radiation, borrowing and dialect continua in the Bantu languages. In Peter Forster and Colin Renfrew (eds.), *Phylogenetic Methods and the Prehistory of Languages*, 19–31. Cambridge: The MacDonald Institute for Archaeological Research.

Holden, Claire J., Andrew Meade, and Mark Pagel. 2005. Comparison of maximum parsimony and Bayesian Bantu language trees. In Ruth Mace, Claire J. Holden, and Stephen Shennan (eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*, 53–65. London: UCL Press.

Huffman, Thomas N. 1980. Ceramics, classifications and Iron Age entities. *African Studies* 29: 123–174.

Huffman, Thomas N. and Robert K. Herbert. 1994–1995. New perspectives on eastern Bantu. *Azania* 29–30: 27–36.

Hyman, Larry M. 1999. The historical interpretation of vowel harmony in Bantu. In Jean-Marie Hombert and Larry M. Hyman (eds.), *Bantu Historical Linguistics: Theoretical and Empirical Perspectives*, 235–295. Stanford: CSLI Publications.

Hyman, Larry M. 2004. How to become a Kwa verb. *Journal of West African Languages* 30: 69–88.

Jackson, Bruce A., Jamie Lee Wilson, Salwa Kirbah, Sheree S. Sidney, Joshua Rosenberger, Larry Bassie, Joe A. Alie, David C. McLean, W. Timothy Garvey, and Bert Ely. 2005. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *American Journal of Physical Anthropology* 128: 156–163.

Jobling, Mark A., Matthew Hurles, and Chris Tyler-Smith. 2004. *Human Evolutionary Genetics: Origins, Peoples and Disease*. New York: Garland.

Jobling, Mark A. and Chris Tyler-Smith. 2003. The human Y chromosome: An evolutionary marker comes of age. *Nature Reviews Genetics* 4: 598–612.

Kayser, Manfred, Silke Brauer, Gunter Weiss, Wulf Schiefenhovel, Peter Underhill, Peidong Shen, Peter Oefner, Mila Tommaseo-Ponzetta, and Mark Stoneking. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *American Journal of Human Genetics* 72: 281–302.

Kivisild, Toomas, Peidong Shen, Dennis P. Wall, Bao Do, Raphael Sung, Karen Davis, Giuseppe Passarino, Peter A. Underhill, Curt Scharfe, Antonio Torroni, Rosaria Scozzari, David Modiano, Alfredo Coppa, Peter de Knijff, Marcus Feldman, Luca L. Cavalli-Sforza, and Peter J. Oefner. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172: 373–387.

Knight, Alec, Peter A. Underhill, Holly M. Mortensen, Lev A. Zhivotovsky, Alice A. Lin, Brenna M. Henn, Dorothy Louis, Merritt Ruhlen, and Joanna L. Mountain. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Current Biology* 13: 464–473.

Lao, Oscar, Timothy T. Lu, Michael Nothnagel, Olaf Junge, Sandra Freitag-Wolf, Amke Caliebe, Miroslava Balascakova, Jaume Bertranpetit, Laurence A. Bindoff, David Comas, Gunilla Holmlund, Anastasia Kouvatsi, Milan Macek, Isabelle Mollet, Walther Parson, Jukka Palo, Rafal Ploski, Antti Sajantila, Adriano Tagliabracci, Ulrik Gether, Thomas Werge, Fernando Rivadeneira, Albert Hofman, André G. Uitterlinden, Christian Gieger, Heinz-Erich Wichmann, Andreas Rüther, Stefan Schreiber, Christian Becker, Peter Nürnberg, Matthew R. Nelson, Michael Krawczak, and Manfred Kayser. 2008. Correlation between genetic and geographic structure in Europe. *Current Biology* 26: 1241–1248.

Lavachery, Philippe. 2001. The Holocene archaeological sequence of Shum Laka rock shelter (Grassfields, Cameroon). *African Archaeological Review* 18: 213–247.

Lwanga-Lunyiigo, Samwiri. 1976. The Bantu problem reconsidered. *Current Anthropology* 17: 282–286.

MacEachern, Scott. 2000. Genes, tribes, and African history. *Current Anthropology* 41: 357–371.

Maho, Jouni. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 639–651. London: Routledge.

Mann, Michael and David Dalby. 1987. *A Thesaurus of African Languages: A Classified and Annotated Inventory of the Spoken Languages of Africa*. London: Hans Zell Publishers.

Mateu, E., D. Comas, F. Calafell, A. Perez-Lezaun, A. Abade, and J. Bertranpetit. 1997. A tale of two islands: Population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. *Annals of Human Genetics* 61: 507–518.

Möhlig, Wilhelm J.G. 1977. Zur frühen Siedlungsgeschichte der Savannen-Bantu aus laut-historischer Sicht. In Wilhelm J.G. Möhlig, Franz Rottland, and Bernd Heine (eds.), *Zur Sprachgeschichte und Ethnohistorie in Afrika*, 166–193. Berlin: Dietrich Reimer.

Möhlig, Wilhelm J.G. 1981. Stratification in the history of the Bantu languages. *Sprache und Geschichte in Afrika* 3: 251–317.

Möhlig, Wilhelm J.G. 1989. Sprachgeschichte, Kulturgeschichte und Archäologie. Die Kongruenz der Forschungsergebnisse als methodologisches Problem. *Paideuma* 35: 189–196.

Neumann, Katharina. 2005. The romance of farming: Plant cultivation and domestication in Africa. In Ann Brower Stahl (ed.), *African Archaeology: A Critical Introduction*, 249–275. Malden: Blackwell.

Novembre, John, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King, Sven Bergmann, Matthew R. Nelson, Matthew Stephens, and Carlos D. Bustamante. 2008. Genes mirror geography within Europe. *Nature* 456: 98–101.

Nurse, Derek. 1997. The contributions of linguistics to the study of history in Africa. *Journal of African History* 38: 359–391.

Nurse, Derek and Gérard Philippson. 2003a. Introduction. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 1–13. London: Routledge.

Nurse, Derek and Gérard Philippson. 2003b. Towards a historical classification of the Bantu languages. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 164–181. London: Routledge.

Oliver, Roland. 1966. The problem of the Bantu expansion. *Journal of African History* 7: 361–376.

Oota, Hiroki, Wannapa Settheetham-Ishida, Danai Tiwawech, Takafumi Ishida, and Mark Stoneking. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature Genetics* 29: 20–21.

Pakendorf, Brigitte. 2007. *Contact in the Prehistory of the Sakha (Yakuts): Linguistic and Genetic Perspectives*. PhD dissertation, Universiteit Utrecht. Downloadable at http://pubman.mpdl.mpg.de/pubman/item/escidoc:405668:4/component/escidoc:405667/sakha_pakendorf2007_s.pdf (accessed January 15, 2011).

Pakendorf, Brigitte, Innokentiy N. Novgorodov, Vladimir L. Osakovskij, and Mark Stoneking. 2007. Mating patterns amongst Siberian reindeer herders: Inferences from mtDNA and Y-chromosomal analyses. *American Journal of Physical Anthropology* 133: 1013–1027.

Pakendorf, Brigitte and Mark Stoneking. 2005. Mitochondrial DNA and human evolution. *Annual Review of Genomics and Human Genetics* 6: 165–183.

Pereira, L., L. Gusmao, C. Alves, A. Amorim, and M.J. Prata. 2002. Bantu and European Y-lineages in Sub-Saharan Africa. *Annals of Human Genetics* 66: 369–378.

Pereira, L., V. Macaulay, A. Torroni, R. Scozzari, M.J. Prata, and A. Amorim. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: Insights into the Bantu expansions and the slave trade. *Annals of Human Genetics* 65: 439–458.

Phillipson, David W. 1977. The spread of the Bantu languages. *Scientific American* 236: 106–114.

Phillipson, David W. 1985. An archaeological reconsideration of Bantu expansion. *Muntu* 2: 69–84.

Phillipson, David W. 1989. The first South African pastoralists and the Early Iron Age. *Nsi* 6: 127–134.

Pilkington, Maya Metni, Jason A. Wilder, Fernando L. Mendez, Murray P. Cox, August Woerner, Thiep Angui, Sarah Kingan, Zahra Mobasher, Chiara Batini, Giovanni Destro-Bisol, Himla Soodyall, Beverly I. Strassmann, and Michael F. Hammer. 2008. Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. *Molecular Biology and Evolution* 25: 517–525.

Plaza, Stéphanie, Antonio Salas, Francesc Calafell, Francisco Corte-Real, Jaume Bertranpetit, Ángel Carracedo, and David Comas. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human Genetics* 115: 439–447.

Pritchard, Jonathan K., Mark T. Seielstad, Anna Perez-Lezaun, and Marcus W. Feldman. 1999. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Molecular Biology and Evolution* 16: 1791–1798.

Quintana-Murci, Lluis, Hélène Quach, Christine Harmant, Franceska Luca, Blandine Massonnet, Etienne Patin, Lucas Sica, Patrick Mouguiama-Daouda, David Comas, Shay Tzur, Oleg Balanovsky, Kenneth K. Kidd, Judith R. Kidd, Lolke van der Veen, Jean-Marie Hombert, Antoine Gessain, Paul Verdu, Alain Froment, Serge Bahuchet, Evelyne Heyer, Jean Dausset, Antonio Salas, and Doron M. Behar. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences of the U.S.A.* 105: 1596–1601.

Quintana-Murci, Lluis, Christine Harmant, Hélène Quach, Oleg Balanovsky, Valery Zaporozhchenko, Connie Bormans, Paul D. van Helden, Eileen G. Hoal, and Doron M. Behar. 2010. Strong maternal Khoisan contribution to the South African coloured population: A case of gender-biased admixture. *American Journal of Human Genetics* 86: 611–620.

Ramachandran, Sohini, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the U.S.A.* 102: 15942–15,947.

Rando, J.C., F. Pinto, A.M. Gonzalez, M. Hernandez, J.M. Larruga, V.M. Cabrera, and H.J. Bandelt. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Annals of Human Genetics* 62: 531–550.

Renfrew, Colin. 1992. World languages and human dispersals: A minimalist view. In John A. Hall and Ian C. Jarvie (eds.), *Transition to Modernity*, 11–68. Cambridge: Cambridge University Press.

Renfrew, Colin. 2002. The ‘emerging synthesis’: The archaeogenetics of farming/language dispersals and other spread zones. In Peter Bellwood and Colin Renfrew (eds.), *Examining the Farming Language Dispersal Hypothesis*, 3–16. Cambridge: McDonald Institute for Archaeological Research.

Rexová, Katerina, Yvonne Bastin, and Daniel Frynta. 2006. Cladistic analysis of Bantu languages: A new tree based on combined lexical and grammatical data. *Naturwissenschaften* 93: 189–194.

Robertson, John H. and Rebecca Bradley. 2000. A new paradigm: The African Early Iron Age without Bantu migrations. *History in Africa* 27: 287–323.

Rosa, Alexandra, António Brehm, Toomas Kivisild, Ene Metspalu, and Richard Villems. 2004. MtDNA profile of West Africa Guineans: Towards a better understanding of the Senegambia region. *Annals of Human Genetics* 68: 340–352.

Rosa, Alexandra, Carolina Ornelas, Mark A. Jobling, António Brehm, and Richard Villems. 2007. Y-chromosomal diversity in the population of Guinea-Bissau: A multiethnic perspective. *BMC Evolutionary Biology* 7: 124.

Salas, Antonio, Martin Richards, Tomás De la Fe, María-Victoria Lareu, Beatriz Sobrino, Paula Sánchez-Diz, Vincent Macaulay, and Ángel Carracedo. 2002. The making of the African mtDNA landscape. *American Journal of Human Genetics* 71: 1082–1111.

Schadeberg, Thilo C. 2003. Historical linguistics. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 143–163. London: Routledge.

Schepartz, Lynne A. 1988. Who were the later Pleistocene Eastern Africans? *African Archaeological Review* 6: 57–72.

Schlebusch, Carina M., Thijessen Naidoo, and Himla Soodyall. 2009. SNaPshot minisequencing to resolve mitochondrial macro-haplogroups found in Africa. *Electrophoresis* 30: 3657–3664.

Scozzari, Rosaria, Fulvio Cruciani, Piero Santolamazza, Patrizia Malaspina, Antonio Torroni, Daniele Sellitto, Barbara Arredi, Giovanni Destro-Bisol, Gianfranco De Stefano, Olga Rickards, Cristina Martinez-Labarga, David Modiano, Gianfranco Biondi, Pedro Moral, Antonel Olckers, Douglas C. Wallace, and Andrea Novelletto. 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *American Journal of Human Genetics* 65: 829–846.

Scozzari, Rosaria, Fulvio Cruciani, Piero Santolamazza, Daniele Sellitto, David E. Cole, Laurence A. Rubin, Damian Labuda, Elisabetta Marini, Valeria Succa, Giuseppe Vona, and Antonio Torroni. 1997. mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: Implications about the origin of their gene pool. *American Journal of Human Genetics* 60: 241–244.

Ségurel, Laure, Begoña Martínez-Cruz, Lluis Quintana-Murci, Patricia Balaresque, Myriam Georges, Tatiana Hegay, Almaz Aldashev, Firuza Nasyrova, Mark A. Jobling, Evelyne Heyer, and Renaud Vitalis. 2008. Sex-specific genetic structure and social organization in Central Asia: Insights from a multi-locus study. *PLoS Genetics* 4: e1000200.

Seielstad, Mark T., Eric Minch, and L. Luca Cavalli-Sforza. 1998. Genetic evidence for a higher female migration rate in humans. *Nature Genetics* 20: 278–280.

Serre, David and Svante Pääbo. 2004. Evidence for gradients of human genetic diversity within and among continents. *Genome Research* 14: 1679–1685.

Sikora, Martin, Hafid Laayouni, Francesc Calafell, David Comas, and Jaume Bertranpetit. 2011. A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations. *European Journal of Human Genetics* 19: 84–88.

Soodyall, Himla, Linda Vigilant, Adrian V. Hill, Mark Stoneking, and Trefor Jenkins. 1996. mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *American Journal of Human Genetics* 58: 595–608.

Templeton, Alan R. 1998. Human races: A genetic and evolutionary perspective. *American Anthropologist* 100: 632–650.

Thomas, Mark G., Tudor Parfitt, Deborah A. Weiss, Karl Skorecki, James F. Wilson, Magdel le Roux, Neil Bradman, and David B. Goldstein. 2000. Y chromosomes traveling south: The Cohen modal haplotype and the origins of the Lemba—the "Black Jews of Southern Africa." *American Journal of Human Genetics* 66: 674–686.

Tishkoff, Sarah A., Mary Katherine Gonder, Brenna M. Henn, Holly Mortensen, Alec Knight, Christopher Gignoux, Neil Fernandopulle, Godfrey Lema, Thomas B. Nyambo, Uma Ramakrishnan, Floyd A. Reed, and Joanna L. Mountain. 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution* 24: 2180–2195.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, Agnes A. Awomoyi, Jean-Marie Bodo, Ogobara Doumbo, Muntaser Ibrahim, Abdalla T. Juma, Maritha J. Kotze, Godfrey Lema, Jason H. Moore, Holly Mortensen, Thomas B. Nyambo, Sabah A. Omar, Kweli Powell, Gideon S. Pretorius, Michael W. Smith, Mahamadou A. Thera, Charles Wambebe, James L. Weber, and Scott M. Williams. 2009. The genetic structure and history of Africans and African Americans. *Science* 324:1035–1044.

Underhill, Peter A., Peidong Shen, Alice A. Lin, Li Jin, Giuseppe Passarino, Wei H. Yang, Erin Kauffman, Batsheva Bonné-Tamir, Jaume Bertranpetit, Paolo Francalacci, Muntaser Ibrahim, Trefor Jenkins, Judith R. Kidd, S. Qasim Mehdi, Mark T. Seielstad, R. Spencer Wells, Alberto Piazza, Ronald W. Davis, Marcus W. Feldman, L. Luca Cavalli-Sforza, and Peter J. Oefner. 2000. Y chromosome sequence variation and the history of human populations. *Nature Genetics* 26: 358–361.

Vansina, Jan. 1979. Bantu in the crystal ball, I. *History in Africa* 6: 287–333.

Vansina, Jan. 1990. *Paths in the Rainforests: Toward a History of Political Tradition in Equatorial Africa*. Madison: University of Wisconsin Press.

Vansina, Jan. 1995. New linguistic evidence and the Bantu expansion. *Journal of African History* 36: 173–195.

Vansina, Jan. 2006. Linguistic evidence for the introduction of ironworking in Bantu-speaking Africa. *History in Africa* 33: 321–361.

Veeramah, Krishna R., Bruce A. Connell, Naser Ansari Pour, Adam Powell, Christopher A. Plaster, David Zeitlyn, Nancy R. Mendell, Michael E. Weale, Neil Bradman, and Mark G. Thomas. 2010. Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evolutionary Biology* 10: 92.

Vigilant, Linda, Mark Stoneking, Henry Harpending, Kristen Hawkes, and Allan C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503–1507.

Watson, Elizabeth, Karin Bauer, Rashid Aman, Gunter Weiss, Arndt von Haeseler, and Svante Pääbo. 1996. mtDNA sequence diversity in Africa. *American Journal of Human Genetics* 59: 437–444.

Wiesmüller, Birgitt. 1997. Möglichkeiten der interdisziplinären Zusammenarbeit von Archäologie und Linguistik am Beispiel der frühen Eisenzeit in Afrika. In Reinhard Klein-Arendt (ed.), *Traditionelles Eisenhandwerk in Afrika: Geschichtliche Rolle und wirtschaftliche Bedeutung aus multidisziplinärer Sicht*, 55–90. Köln: Heinrich-Barth-Institut für Archäologie und Geschichte Afrikas.

Williamson, Kay and Roger Blench. 2000. Niger-Congo. In Bernd Heine and Derek Nurse (eds.), *African Languages: An Introduction*, 11–42. Cambridge: Cambridge University Press.

Wood, Elizabeth T., Daryn A. Stover, Christopher Ehret, Giovanni Destro-Bisol, Gabriella Spedini, Howard McLeod, Leslie Louie, Mike Bamshad, Beverly I. Strassmann, Himla Soodyall, and Michael F. Hammer. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: Evidence for sex-biased demographic processes. *European Journal of Human Genetics* 13: 867–876.

Zhivotovsky, Lev A., Peter A. Underhill, Cengiz Cinnioğlu, Manfred Kayser, Bharti Morar, Toomas Kivisild, Rosaria Scozzari, Fulvio Cruciani, Giovanni Destro-Bisol, Gabriella Spedini, Geoffrey K. Chambers, Rene J. Herrera, Kiau Kiun Yong, David Gresham, Ivailo Tournev, Marcus W. Feldman, and Luba Kalaydjieva. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *American Journal of Human Genetics* 74: 50–61.

Appendix. Ethnolinguistically defined Niger-Congo populations included in molecular anthropological studies to date, restricted to the uniparental markers (mtDNA and Y chromosome)

While this table is as accurate as possible, it is not always easy to trace samples through several publications to their original source; therefore, some errors may have slipped in, especially with respect to the sample numbers. Not all the studies referenced here as having included African samples deal primarily with the prehistory of African populations; in some cases, samples from African ethnolinguistic groups were included in studies of worldwide variation. On the other hand, some studies, such as the investigation of genetic variation in the Angolan enclave of Cabinda (Beleza et al., 2005), do not provide ethnolinguistic information on the samples included; these have been excluded from the table.

Populations with sample sizes smaller than 15 are omitted, with a few exceptions where the same population has more than 14 individuals typed either for the mtDNA or Y chromosome. In the column 'subgroup,' the language code according to Guthrie's (1971) and Maho's (2003) classification of Bantu languages is given in parentheses, while the column 'ID' contains the code used in Fig. 3. The column labeled 'N' gives the sample sizes for studies of mtDNA and Y chromosome, respectively. The references are included in the bibliography at the end of the paper.

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Baka Pygmies	Adamawa	bkP	127	Quintana-Murci et al., 2008	56	Wood et al., 2005; Berniell-Lee et al., 2009
Fali	Adamawa	fal	40	Černý et al., 2007	39	Scozzari et al., 1997; Scozzari et al., 1999
Balanta	Atlantic	bal	62	Rosa et al., 2004	26	Rosa et al., 2007
Borgor Fulani	Atlantic	bfu	49	Černý et al., 2007	–	–
Felupe Ejamat	Atlantic	fel	50	Rosa et al., 2004	50	Rosa et al., 2007
Fulbe	Atlantic	fu1, fu2, fu3	61	Vigilant et al., 1991; Watson et al., 1996; Kivisild et al., 2006	59	Scozzari et al., 1997; Destro-Bisol et al., 2004b; Rosa et al., 2007
Limba	Atlantic	lim	67	Jackson et al., 2005	–	–

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Nalu	Atlantic	nal	26	Rosa et al., 2004	17	Rosa et al., 2007
Papel	Atlantic	pap	77	Rosa et al., 2004	64	Rosa et al., 2007
Serer	Atlantic	ser	23	Rando et al., 1998	–	–
Tcheboua Fulani	Atlantic	tfu	40	Černý et al., 2007	–	–
Temne	Atlantic	tem	121	Jackson et al., 2005	–	–
Wolof	Atlantic	wol	48	Rando et al., 1998	34	Wood et al., 2005
Annang	Cross-River	ann	110	Veeramah et al., 2010	110	Veeramah et al., 2010
Efik	Cross-River	efi	148	Veeramah et al., 2010	148	Veeramah et al., 2010
Ibibio	Cross-River	ibi	516	Veeramah et al., 2010	516	Veeramah et al., 2010
Oron	Cross-River	oro	101	Veeramah et al., 2010	101	Veeramah et al., 2010
Yoruba	Defoid	yor	34	Vigilant et al., 1991; Watson et al., 1996	13	Tishkoff et al., 2007; de Filippo et al., 2011
Dogon	Dogon	dog	–	–	55	Wood et al., 2005
Kassena	Gur	kas	–	–	33	de Filippo et al., 2011
Lyela	Gur	lye	–	–	40	de Filippo et al., 2011
Mossi	Gur	mos	10	Kivisild et al., 2006	92	Scozzari et al., 1997; Scozzari et al., 1999; de Filippo et al., 2011
Nuna	Gur	nun	–	–	29	de Filippo et al., 2011
Pana	Gur	pan	–	–	19	de Filippo et al., 2011
Samoya	Gur	sam	–	–	21	de Filippo et al., 2011
Ewe	Kwa	ewe	88	Veeramah et al., 2010	118	Wood et al., 2005; Veeramah et al., 2010
Fante	Kwa	fan	–	–	32	Wood et al., 2005

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Fon	Kwa	fon	–	–	36	Scozzari et al., 1999
Ga	Kwa	ga	–	–	29	Wood et al., 2005
Akan	Kwa	aka	155	Veeramah et al., 2010	155	Veeramah et al., 2010
Bambara	Mande	bab	52	González et al., 2006	–	–
Bissa	Mande	bis	–	–	40	de Filippo et al., 2011
Loko	Mande	lok	29	Jackson et al., 2005	–	–
Malinke	Mande	mal	31	González et al., 2006	–	–
Mandenka	Mande	mn1, mn2	119	Graven et al., 1995	99	Wood et al., 2005; Rosa et al., 2007; de Filippo et al., 2011
Marka	Mande	mrk	–	–	33	de Filippo et al., 2011
Mende	Mande	men	59	Jackson et al., 2005	–	–
Samo	Mande	smo	–	–	78	de Filippo et al., 2011
Bariba	Voltaic	bar	–	–	34	Scozzari et al., 1999
Berba	Voltaic	brb	–	–	20	Scozzari et al., 1999
Igbo	West Benue-Congo	igb	209	Veeramah et al., 2010	209	Veeramah et al., 2010
Aghem	Bantoïd	agh	118	Veeramah et al., 2010	118	Veeramah et al., 2010
Bamileke	Bantoïd	bam	48	Destro-Bisol et al., 2004a	48	Scozzari et al., 1997; Caglià et al., 2003; Coia et al., 2004
Bamun	Bantoïd	bmn	117	Veeramah et al., 2010	117	Veeramah et al., 2010
Ejagham	Bantoïd	eja	138	Veeramah et al., 2010	117	Veeramah et al., 2010
Tikar	Bantoïd	tik	34	Veeramah et al., 2010	34	Veeramah et al., 2010
Tikar Pygmies	Bantoïd	tiP	35	Quintana-Murci et al., 2008	–	–

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Bubi	Bantu (A31)	A31	45	Mateu et al., 1997	–	–
Benga	Bantu (A34)	A34	50	Quintana-Murci et al., 2008	48	Berniell-Lee et al., 2009
Bassa	Bantu (A43)	A43	47	Destro-Bisol et al., 2004b	49	Destro-Bisol et al., 2004b
Ewondo	Bantu (A72)	A72	53	Destro-Bisol et al., 2004a; Quintana-Murci et al., 2008	30	Scozzari et al., 1997; Caglià et al., 2003; Coia et al., 2004
Fang	Bantu (A75)	A75	105	Quintana-Murci et al., 2008	64	Berniell-Lee et al., 2009
Makina	Bantu (A803)	A803	45	Quintana-Murci et al., 2008	43	Berniell-Lee et al., 2009
Ngoumba	Bantu (A81)	A81	88	Quintana-Murci et al., 2008	55	Wood et al., 2005; Berniell-Lee et al., 2009
Bakaka	Bantu (A93)	A93	50	Destro-Bisol et al., 2004b	49	Destro-Bisol et al., 2004b
Galoa	Bantu (B11)	B11	51	Quintana-Murci et al., 2008	47	Berniell-Lee et al., 2009
Orungu	Bantu (B11)	B11	20	Quintana-Murci et al., 2008	21	Berniell-Lee et al., 2009
Akele	Bantu (B22)	B22	48	Quintana-Murci et al., 2008	50	Berniell-Lee et al., 2009
Kota	Bantu (B25)	B25	56	Quintana-Murci et al., 2008	53	Berniell-Lee et al., 2009
Shake	Bantu (B251)	B251	51	Quintana-Murci et al., 2008	43	Berniell-Lee et al., 2009
Eviya	Bantu (B301)	B301	38	Quintana-Murci et al., 2008	24	Berniell-Lee et al., 2009
Mitsogo	Bantu (B31)	B31	64	Quintana-Murci et al., 2008	60	Berniell-Lee et al., 2009
Eshira	Bantu (B41)	B41	40	Quintana-Murci et al., 2008	42	Berniell-Lee et al., 2009
Punu	Bantu (B43)	B43	52	Quintana-Murci et al., 2008	58	Berniell-Lee et al., 2009

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Duma	Bantu (B51)	B51	47	Quintana-Murci et al., 2008	46	Berniell-Lee et al., 2009
Nzebi	Bantu (B52)	B52	63	Quintana-Murci et al., 2008	57	Berniell-Lee et al., 2009
Obamba	Bantu (B62)	B62	47	Quintana-Murci et al., 2008	47	Berniell-Lee et al., 2009
Ndumu	Bantu (B63)	B63	39	Quintana-Murci et al., 2008	36	Berniell-Lee et al., 2009
Ateke	Bantu (B71)	B71	54	Quintana-Murci et al., 2008	48	Berniell-Lee et al., 2009
Yansi	Bantu (B85)	B85	–	–	23	de Filippo et al., 2011
Nande	Bantu (D/J42)	D42	–	–	18	Wood et al., 2005
Hutu	Bantu (D/J61)	D61	42	Castrì et al., 2009	39	Caglià et al., 2003
Hema	Bantu (E/J10)	E10	–	–	18	Wood et al., 2005
Ganda	Bantu (E/J15)	E15	–	–	26	Wood et al., 2005
Kikuyu	Bantu (E51)	E51	25	Watson et al., 1996	42*	Wood et al., 2005
Sukuma	Bantu (F21)	F21	32	Knight et al., 2003; Tishkoff et al., 2007	32	Knight et al., 2003; Tishkoff et al., 2007
Turu	Bantu (F32)	F32	29	Tishkoff et al., 2007	20	Tishkoff et al., 2007
Mbugwe	Bantu (F34)	F34	–	–	15	Tishkoff et al., 2007
Luvale	Bantu (K14)	K14	–	–	16	de Filippo et al., 2011
Mbunda	Bantu (K15)	K15	–	–	49	de Filippo et al., 2011
Ganguela	Bantu (K19)	K19	21	Coelho et al., 2009	21	Coelho et al., 2009
Luyana	Bantu (K31)	K31	–	–	61**	de Filippo et al., 2011

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Kwamashi	Bantu (K34)	K34	–	–	27	de Filippo et al., 2011
Shanjo	Bantu (K36)	K36	–	–	29	de Filippo et al., 2011
Fwe	Bantu (K402)	K402	–	–	30	de Filippo et al., 2011
Totela	Bantu (K41)	K41	–	–	15	de Filippo et al., 2011
Nkoya	Bantu (L62)	L62	–	–	15	de Filippo et al., 2011
Bisa	Bantu (M51)	M51	42	de Filippo et al., 2010	36	de Filippo et al., 2010; 2011
Tonga	Bantu (M64)	M64	–	–	30	de Filippo et al., 2011
Nyanja	Bantu (N31)	N31	20	Salas et al., 2002	–	–
Kunda	Bantu (N42)	N42	36	de Filippo et al., 2010	36	de Filippo et al., 2010; 2011
Nyungwe	Bantu (N43)	N43	20	Salas et al., 2002	–	–
Sena	Bantu (N44)	N44	21	Salas et al., 2002	–	–
Makonde	Bantu (P23)	P23	19	Salas et al., 2002	–	–
Makhuwa	Bantu (P31)	P31	20	Salas et al., 2002	–	–
Lomwe	Bantu (P32)	P32	20	Salas et al., 2002	–	–
Chwabo	Bantu (P34)	P34	20	Salas et al., 2002	–	–
Mbundu	Bantu (probably H21)	H21	44	Plaza et al., 2004	–	–
Umbundu	Bantu (R11)	R11	96	Coelho et al., 2009	96	Coelho et al., 2009
Nkhumbi***	Bantu (R14)	R14	153	Coelho et al., 2009	153	Coelho et al., 2009

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Ovambos	Bantu (R21)	R21	–	–	28	Forster et al., 2000
Ambo	Bantu (R22)	R22	–	–	22	Wood et al., 2005
Herero	Bantu (R31)	R31	–	–	24	Wood et al., 2005
Kuvale	Bantu (R31)	R31	56	Coelho et al., 2009	56	Coelho et al., 2009
Shona	Bantu (S10)	S10	77	Salas et al., 2002; Castrì et al., 2009	49	Wood et al., 2005
Ndau	Bantu (S15)	S15	19	Salas et al., 2002	–	–
Kalanga	Bantu (S16)	S16	–	–	20	de Filippo et al., 2011
Sotho-Tswana	Bantu (S31–32)	S30	–	–	28	Caglià et al., 2003; Wood et al., 2005
Tswana	Bantu (S31)	S31	–	–	20	de Filippo et al., 2011
Lozi	Bantu (S34)	S34	–	–	94	de Filippo et al., 2011
Xhosa	Bantu (S41)	S41	–	–	80	Wood et al., 2005
Zulu	Bantu (S42)	S42	–	–	29	Wood et al., 2005
Tswa	Bantu (S51)	S51	19	Salas et al., 2002	–	–
Shangaan	Bantu (S53)	S53	22	Salas et al., 2002	25	Pereira et al., 2002
Ronga	Bantu (S54)	S54	21	Salas et al., 2002	15	Pereira et al., 2002
Chopi	Bantu (S61)	S61	27	Salas et al., 2002	–	–
Mbenzele Pygmies	Bantu (C10)	mbP	57	Destro-Bisol et al., 2004a	52	Coia et al., 2004
Babongo Pygmies	Bantu (B42)	B42	45	Quintana-Murci et al., 2008	–	–

			Mitochondrial DNA		Y chromosome	
Population	subgroup	ID	N	Study	N	Study
Bakola Pygmies	Bantu (A801)	A801	88	Tishkoff et al., 2007; Quintana-Murci et al., 2008	55	Wood et al., 2005; Berniell-Lee et al., 2009
Bakoya Pygmies	Bantu (B22b)	B22b	31	Quintana-Murci et al., 2008	–	–
Biaka Pygmies	Bantu (C104)	C104	56	Vigilant et al., 1991; Kivisild et al., 2006; Quintana-Murci et al., 2008	54	Pritchard et al., 1999; Underhill et al., 2000; Cruciani et al., 2002; Destro-Bisol et al., 2004b; Wood et al., 2005; de Filippo et al., 2011

* a mixed sample of Kikuyu and Kamba

** a mixed sample of Kwangwa, Makoma, Kwandi, Kwamulonga, Mwenyi, Mbowe, Simaa, and Luyana

*** a mixed sample of Nyaneka and Nkhumbi