Static and Moving Frontiers: The Genetic Landscape of Southern African Bantu-Speaking Populations

Sarah J. Marks,[†,1] Francesco Montinaro,[†,1,2] Hila Levy,[1] Francesca Brisighelli,[2] Gianmarco Ferri,[3] Stefania Bertoncini,[4] Chiara Batini,[5] George B.J. Busby,[1] Charles Arthur,[6] Peter Mitchell,[6,7] Brian A. Stewart,[8] Ockie Oosthuizen,[9] Erica Oosthuizen,[9] Maria Eugenia D'Amato,[10] Sean Davison,[10] Vincenzo Pascali,[2] and Cristian Capelli*[,1]

[1]Department of Zoology, University of Oxford, Oxford, United Kingdom

[2]Institute of Legal Medicine, Catholic University, Rome, Italy

[3]Dipartimento ad Attività Integrata di Laboratori, Anatomia Patologica, Medicina Legale, U.O. Struttura Complessa di Medicina Legale, Azienda Ospedaliero, Universitaria di Modena, Modena, Italy

[4]Department of Biology, University of Pisa, Pisa, Italy

[5]Department of Genetics, University of Leicester, Leicester, United Kingdom

[6]School of Archaeology, University of Oxford, Oxford, United Kingdom

[7]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

[8]Museum of Anthropology, University of Michigan

[9]School of Medicine, University of Namibia, Windhoek, Namibia

[10]Biotechnology Department, Forensic DNA Laboratory, University of the Western Cape, Bellville, South Africa

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: cristian.capelli@zoo.ox.ac.uk.

Associate editor: Sarah Tishkoff

Abstract

A consensus on Bantu-speaking populations being genetically similar has emerged in the last few years, but the demographic scenarios associated with their dispersal are still a matter of debate. The frontier model proposed by archeologists postulates different degrees of interaction among incoming agropastoralist and resident foraging groups in the presence of "static" and "moving" frontiers. By combining mitochondrial DNA and Y chromosome data collected from several southern African populations, we show that Bantu-speaking populations from regions characterized by a moving frontier developing after a long-term static frontier have larger hunter-gatherer contributions than groups from areas where a static frontier was not followed by further spatial expansion. Differences in the female and male components suggest that the process of assimilation of the long-term resident groups into agropastoralist societies was gender biased. Our results show that the diffusion of Bantu languages and culture in Southern Africa was a process more complex than previously described and suggest that the admixture dynamics between farmers and foragers played an important role in shaping the current patterns of genetic diversity.

Key words: **"Bantu expansion", Y chromosome, mtDNA, frontier model, admixture.**

Introduction

Migrations and subsequent admixture between populations have shaped the worldwide distribution of genetic variation (Pickrell and Pritchard 2012; Pickrell et al. 2012, 2014; Hellenthal et al. 2014). Within sub-Saharan Africa, the expansion of iron-using agropastoralist populations speaking Bantu languages has commonly been described as the most influential demographic event to have occurred on the African continent (Diamond 1997). This dispersal is believed to have started around 5,000 years ago, with movement from northwestern Cameroon/southern Nigeria throughout most of Africa south of the Equator, the southernmost regions being among the last to be occupied (Newman 1995; de Filippo et al. 2012). The arrival of Bantu-speaking farmers potentially led to the isolation of, and/or admixture with, the hunter-gatherer and pastoralist groups already present in these areas, including Pygmies in Central Africa and the Khoekhoe and San groups of Southern Africa (Destro-Bisol et al. 2004; Mitchell et al. 2008; Mitchell 2009; Patin et al. 2014).

A working framework that might help to explain and predict the structure of genetic variation resulting from this dispersal is offered by the "frontier" model. Frontier dynamics were first proposed in the late 19th and early 20th centuries by researchers studying the spread of European farmers in North America, Asia, and Australia (Burt 1940; Turner 1962; Alexander 1977). The term frontier has been defined as "the temporary boundary of an expanding society at the edge of substantially free lands" (Turner 1962) and several such frontiers have been identified worldwide. Alexander (1977) proposed the presence of two main types of frontiers when considered from the point of view of the farmers: "moving" and "static." Moving frontiers characterize the expansion of farming societies and/or their technologies into new regions in the absence of ecological and geographical restrictions. In such a situation, after an initial pioneer phase, the farmers "subdue the wilderness," as opposed to only exploiting some aspects of it, increase their population size, and continue to expand into new areas (Ammerman and Cavalli-Sforza 1973; Alexander 1984). At this stage, the interactions between groups with different subsistence strategies are limited, with hunter-gatherers attempting to retain autonomy by retreating into more isolated/agriculturally less favorable areas and farmers seizing new land to sustain their dispersal. Once usable land is exhausted, natural boundaries (e.g., seas or mountains) are reached, or farmers' crops and animals cannot tolerate the ecological conditions encountered, further dispersal is prevented. Moving frontiers then become static: Changes occur in farmers' social organization, often resulting in long-term relationships with neighboring hunter-gatherers if still present (Bohannan and Plog 1967; Alexander 1984). As interaction increases and farmers' numbers grow, hunter-gatherer autonomy is curtailed and the attractiveness of assimilation into farming communities enhanced to the point where no other opportunities exist (Wadley 1996): The latter's absorption often being favored by their ability to offer specialized skills and services (Hammond-Tooke 1998, 1999).

Such a model implies limited assimilation of foragers into farmers' communities during the moving frontier phase of the dispersal process, with an increasing likelihood of gene flow with foragers once the process of local colonization and population expansion has concluded and a static frontier emerges. Static frontiers are expected to develop after moving ones; however, the signature of hunter-gatherer assimilation into farming communities is shaped by the degree of survival of the former and the population size of the latter. The development of a static frontier coincides with the farmers' population reaching density close to the carrying capacity of the area, preventing further significant local population expansion. Simulations have shown that gene flow between incoming and resident groups that might occur at this stage is expected to leave a marginal signature unless subsequent expansion of the admixed group occurs by moving to new areas (Currat et al. 2008).

Alexander (1977, 1984) employed ecological and archeological data to argue for the presence of static and moving frontiers in Southern Africa associated with the dispersal of Bantu-speaking farming populations. In Malawi, Mozambique, Zambia, and Zimbabwe, here defined as southcentral (SC) Africa, farmers moved across and occupied most of the available space relatively quickly, reaching a relatively high population density and entering the conditions of a static frontier. To the south of the Limpopo River (Lesotho and eastern part of South Africa) and also westward across the southern part of Botswana (the two areas here grouped together as southeastern Africa [SE]) and Namibia, possibilities of expansion were stalled by an ecoclimatic barrier running from the Eastern Cape Province, along the foothills of the uKhahlamba-Drakensberg escarpment and then north along the edge of the Grassland Biome and the eastern and northern borders of the Kalahari Desert (fig. 1). Major constraints to their expansion beyond this line in the first millennium AD were set by their reliance on summer rainfall cereals (sorghum, pearl millet, and finger millet). Subsequently, farmers moved into the grassland biome of upland KwaZulu Natal, Swaziland, and the central interior of South Africa from around AD 1300, a process that gathered pace around the mid-1600s. Finally, the dispersal of maize (introduced by the Portuguese Mozambique in the 16th century) and temporarily wetter conditions first allowed a temporary further expansion of agropastoralist communities into drier areas of Southern Africa's summer rainfall zone in the 18th century (Huffman 1988) and then (along with wheat and other crops of European origin) permitted the permanent agropastoralist colonization of Lesotho's Maloti Mountains during the 1800s (Gill 1993).

Archeological data suggest that hunter-gatherers had disappeared across much of Zambia, Zimbabwe, Malawi, and Mozambique by the early second millennium AD (Walker 1995; Phillipson 2005). Further south, a similar pattern is evident across the eastern part of Southern Africa, where several studies indicate that hunter-gatherers may have engaged in relatively equitable exchange relations with incoming farmers (Mazel 1989; Mitchell 2009). However, as agropastoralist populations grew and expanded in the second millennium AD, these relationships became increasingly unequal, leading to the eventual incorporation and assimilation of foraging groups. Decreasing forager access to key resources or parts of the landscape and loss of hunter-gatherer women through hypergamic marriage are thought to have been particularly important in these developments (Mazel 1989; Wadley 1996; Hammond-Tooke 1998, 1999). Today, no Khoesan-speaking groups remain in Lesotho and the adjacent provinces of South Africa where people speak one or more Bantu languages: Sesotho, Sephuti, isiZulu, and isiXhosa.

According to the frontier model, the different admixture and demographic dynamics that characterized the dispersal of farmers across SC and SE Africa shaped the genetic composition of Bantu-speaking populations from these areas, in particular in relation to the amount of hunter-gatherer contribution, which is expected to be higher in groups from the SE than the SC region. In order to test if Bantu-speaking groups from these two areas are, indeed, characterized by different genetic profiles as the result of differential admixture between farmers and hunter-gatherers, we investigated the mitochondrial DNA (mtDNA) and Y chromosome variation of populations from Lesotho (Sotho, Ndebele, and Thembu) and nearby regions of South Africa (Xhosa and Zulu populations). By combining novel data with previously published results, we investigated the contribution of the frontier model in explaining the distribution of genetic variation in SC and SE Africa, in an attempt to shed light on the demographic dynamics associated with the expansion of Bantu-speaking agropastoralists south of the Equator.

Results

Genetic Diversity

The mtDNA and Y chromosome haplogroup composition of the newly reported populations is presented in supplementary table S2, Supplementary Material online. For the mtDNA, all the Bantu-speaking populations displayed relatively high haplotype diversity (HD) spanning from 0.969 to 0.983 (table 1). Bantu-speaking populations also show significant Fu' Fs values, possibly linked to the demographic growth related to their dispersal. For the Y chromosome, the Thembu population (characterized by the smallest sample size) shows a low HD value but a relatively large standard deviation (0.938 ± 0.041). At the haplogroup level (supplementary table S2*a*, Supplementary Material online), most of the mitochondrial haplotypes found in the Lesotho and South African populations belong to the sub-Saharan African haplogroup *L*(*xM*, *N*). A large proportion of these (32.5%) fall within L0d, a haplogroup found at high frequencies in Khoesan populations (Vigilant et al. 1991; Chen et al. 2000; Salas et al. 2002; Kivisild et al. 2004; Barbieri, Vicente, et al. 2013; Schlebusch et al. 2013).

Similarly to mtDNA, a Y haplogroup commonly present in Khoesan populations was also observed in the paternal gene pool of the Bantu-speaking populations presented here (haplogroup A3b1-M51; Knight et al. 2003; Tishkoff et al. 2007; Batini et al. 2011), ranging from 1.8% in the Xhosa to 11% in the Sotho (supplementary table S2*b*, Supplementary Material online). The majority of the SE African Y chromosomes analyzed here (supplementary table S2b, Supplementary Material online) fall into E-M96, a pan-African haplogroup associated with the Bantu expansion, which is often of high frequency in Bantu-speaking populations (Coelho et al. 2009; Quintana-Murci et al. 2010; de Filippo et al. 2011). Another Bantu-associated haplogroup, B2a-M150 (Gomes et al. 2010; but see also Batini et al. 2011), is also found in the Southern African samples, at an average frequency of 9.9% across the five Bantu-speaking populations.

The Ju/'hoãnsi San haplogroup composition is very similar to that previously reported for related populations (Cruciani et al. 2002; Wood et al. 2005; Tishkoff et al. 2007; Barbieri, Vicente, et al. 2013; Schlebusch et al. 2013; Barbieri et al. 2014). mtDNA haplogroups all belong to L0d and L0k haplotypes, almost uniquely found only in Southern Africa. Y chromosome haplotypes were represented by A2, A3b1, and B2b haplogroups, with few haplotypes belonging to haplogroup E.

The Ju/'hoãnsi San was characterized by lower HD values than Bantu-speaking groups for mtDNA but not the Y chromosome, while Fu's Fs was nonsignificant. The Y chromosome (but not mtDNA) MNPD (mean number of pairwise differences) value was larger than those in the other populations presented here.

Population Structure

In order to explore the degree of differentiation and population structure among SE and SC African Bantu-speaking populations, we assembled a data set comprising several populations from this area (fig. 1 and supplementary table S1, Supplementary Material online). We initially investigated the correlation between genetic and geographic distances by a Mantel test (Mantel 1967; Mantel and Valand 1970). A significant correlation was observed for mtDNA ($r = 0.353$, $P < 0.001$) and the Y chromosome ($r = 0.56$, $P < 0.001$), as previously reported (de Filippo et al. 2012).

The degree of differentiation among the populations was investigated by Multidimensional Scaling (MDS) analysis based on Reynolds' distance calculated using haplotype frequencies (Reynolds et al. 1983). When analyzed for the mtDNA that the distribution of populations in the plot supports the correlations between geographic and genetic distances, with populations from the same country clustering together (fig. 2*a*). However, we also observe that in some cases populations from distant sampling locations clustered closely in the plot. This is particularly evident in the lower part of the MDS plot, where most of the populations from SE Africa (Lesotho, South Africa, and Botswana, within Area 2 in fig. 1) and a few populations from Angola and Zambia group together along the second dimension. This observation suggests that other factors beside their common origin might have influenced the genetic structure of Bantu-speaking groups. Intrigued by the detection of L0d/L0K haplogroups in Bantu-speaking groups, we tested for the correlation between the genetic distance of each population and the forager Ju/'hoãnsi and the position in the MDS plot of each of the populations in the assembled data set and found a significant association (after Bonferroni correction) with Dimension 2 ($R^2 = 0.48$, $P < 0.001$, fig. 2*b*). We then tested the correlation between the amount of L0d/L0k in each population and the positioning along the MDS plot. After Bonferroni correction, significant correlation is present between L0d/L0K frequencies and the populations positioning along Dimension 2 of

Table 1. Within-Population Diversity Summary Statistics for the Populations Presented in This Work.

	mtDNA						Y Chromosome			
Population	*N*	*k*	HD	MNPD	Tajima's *D*	Fu's Fs	*N*	*k*	HD	MNPD
Ndebele	33	25	0.983 ± 0.011	9.795 ± 4.601	−0.567 (ns)	−9.102 ($P < 0.05$)	48	37	0.984 ± 0.009	6.333 ± 3.187
Sotho	287	103	0.969 ± 0.004	9.140 ± 4.218	−0.867 (ns)	−24.149 ($P < 0.05$)	181	123	0.992 ± 0.002	7.399 ± 3.476
Thembu	24	20	0.982 ± 0.018	8.449 ± 4.051	−0.504 (ns)	−8.426 ($P < 0.05$)	21	15	0.938 ± 0.041	5.886 ± 2.928
Xhosa	54	37	0.971 ± 0.012	8.197 ± 3.861	−0.590 (ns)	−20.404 ($P < 0.05$)	57	45	0.990 ± 0.006	4.618 ± 3.606
Zulu	54	32	0.971 ± 0.010	9.134 ± 4.268	−0.302 (ns)	−10.804 ($P < 0.05$)	51	46	0.994 ± 0.007	7.718 ± 3.656
Ju/'hoãnsi	58	16	0.925 ± 0.013	6.478 ± 0.013	0.093 (ns)	−0.118 (ns)	53	34	0.978 ± 0.009	9.622 ± 4.481

N = sample size; *k* = number of haplotypes; ns = nonsignificant.

the MDS plot ($R^2 = 0.49$, $P < 0.001$, fig. 2*c*). This correlation is still observed even when the populations that do not show L0d/L0K haplogroups are discarded ($R^2 = 0.40$, $P < 0.001$; data not shown). Reynolds' distance used to generate the MDS plot is based on haplotype frequencies and does not take into account the molecular distance between haplotypes. For this reason, the MDS positions are not expected to be necessarily correlated with haplogroup frequencies.

We similarly explored the Y chromosome Short Tandem Repeat (STR) haplotype data and generated a population-based MDS plot (fig. 3*a*). Mirroring mtDNA results, a significant correlation between the positioning of populations on the MDS plot (Dimension 1) and the genetic distance between each population and the Ju/'hoãnsi was observed (fig. 3*b*). Moreover, significant correlation was found between the frequency of the Y chromosome haplogroup A (common in foraging communities from Southern Africa; Underhill et al. 2000; Cruciani et al. 2002; Tishkoff et al. 2007; Batini et al. 2011) and the first MDS component (fig. 3*c*, $R^2 = 0.43$, $P < 0.001$). The correlation is still present when the B haplogroup is included ($R^2 = 0.38$, $P < 0.001$).

The relevance of lineages commonly found in foraging communities in shaping the genetic variation of Southern African Bantu-speaking populations is further confirmed by correspondence analyses. Haplogroups L0d and L0K drive the clustering in the upper left and right corners of the mtDNA population plot respectively (supplementary fig. S1, Supplementary Material online). The role played by haplogroup A in shaping the Y chromosome correspondence analysis plot is evident despite the small number of haplogroups analyzed (supplementary fig. S2, Supplementary Material online). Notably, for the mtDNA, all the populations from the SE African region (within Area 2 of fig. 1) cluster together with Ronga from Mozambique and Kuvale from Angola as a result of the occurrence of haplogroup L0d in these populations (supplementary fig. S1, Supplementary

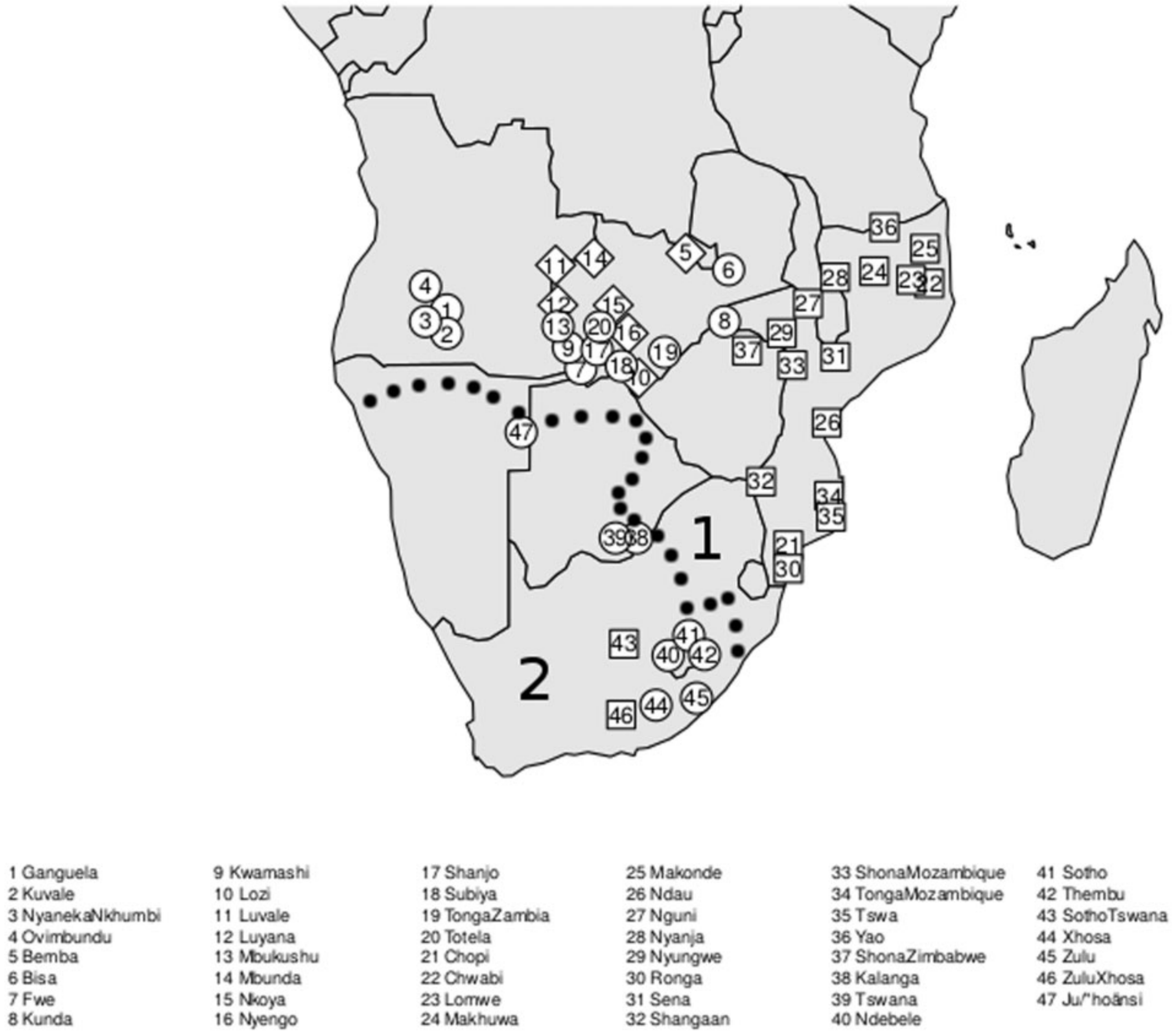

Fig. 1. Map of the analyzed populations. Circles indicate populations for which both mtDNA and Y chromosome data are available; squares and diamonds represent populations for which only mtDNA or Y chromosome data are available, respectively. The dotted line broadly describes the extension of the ecoclimatic boundary running across Southern Africa and the two ecological regions defined by this (indicated by 1 and 2) as described in the text.

Material online). This is also evident in the correspondence analysis of the Y chromosome, where haplogroup A drives the localization of some populations within Areas 1 and 2 (supplementary fig. S2, Supplementary Material online).

We applied a discriminant analysis of principal components (DAPC) to search for evidence of structure among the analyzed populations (Jombart et al. 2010). In this analysis, we used mutation frequencies instead of haplogroups in order to avoid haplogroup misclassification due to incomplete resolution provided by Hypervariable region I (HVR-I) (mtDNA) or differences in the single nucleotide polymorphisms (SNPs) analyzed across different studies (Y chromosome). Mutation frequencies were successfully applied in previous investigations (Montano et al. 2013). According to the Bayesian information criterion (BIC; supplementary fig. S3, Supplementary Material online) distribution, the optimal

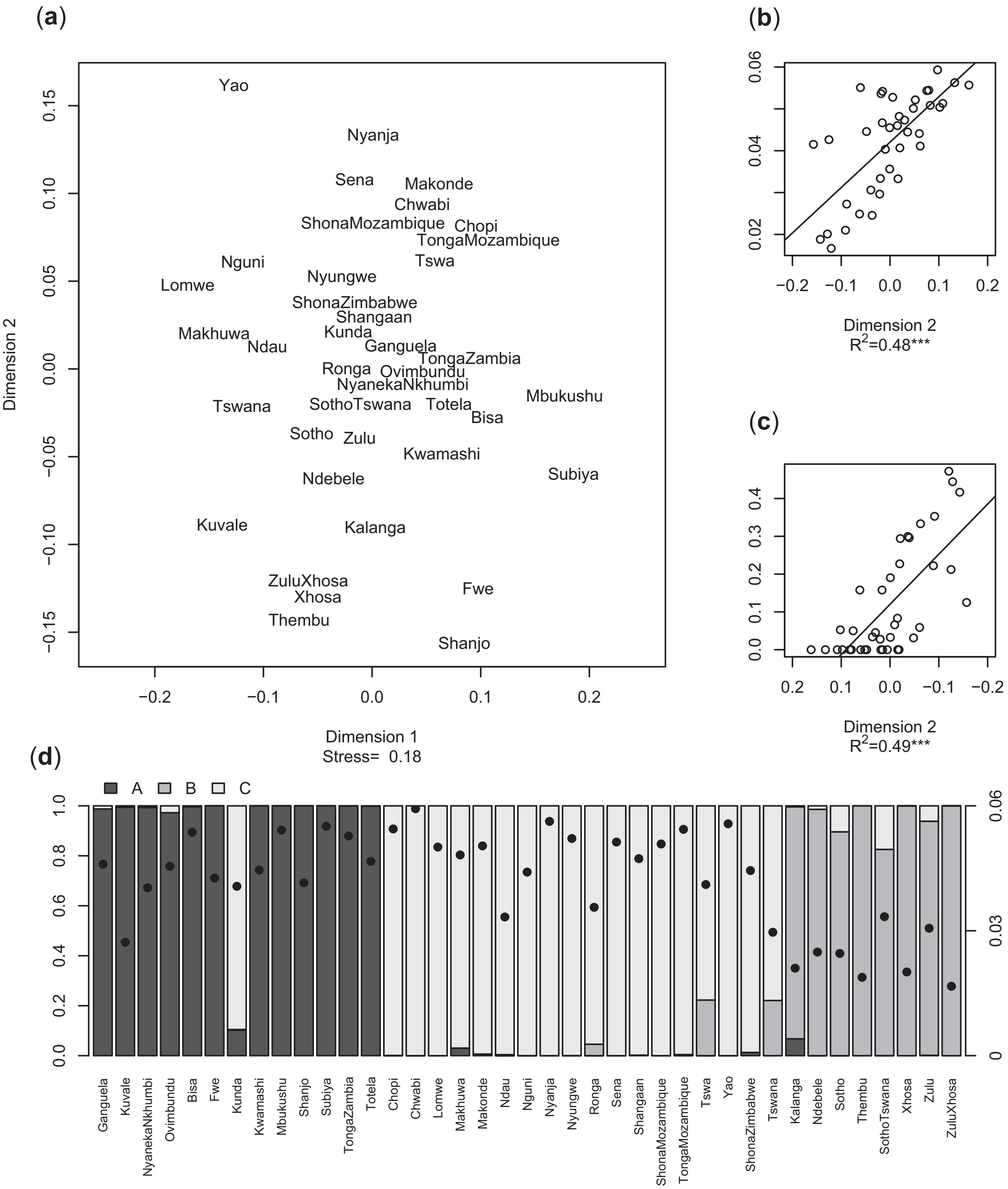

Fig. 2. (*a*) MDS plot based on the Reynolds' distance matrix for mtDNA. (*b*, *c*) Correlation between the second component of the MDS plot in (*a*) versus (*b*) the Reynolds' distance between each populations and the Ju/'hoãnsi and (*c*) the amounts of L0d/L0k mtDNA haplogroups in the analyzed populations. (*d*) Cluster assignation probability as defined by DAPC analysis. Dots within each bar indicate the genetic distance of the indicated populations with the Ju/'hoãnsi.

number of clusters was $K = 3$ when considering mtDNA data (supplementary figs. S3*a* and S4*a*, Supplementary Material online), whereas for the Y chromosome, $K = 2$ was the optimal choice (supplementary figs. S3*b* and S4*b*, Supplementary Material online). Both genetic systems provided support for a cluster almost completely represented by populations from the SE African region (Lesotho, Botswana, and South Africa), with the exception of Tswana and Kalanga for mtDNA and Y chromosome, respectively, additionally including the Subiya from Zambia for the Y chromosome (cluster B in figs. 2*d* and 3*d*). We noted that for both genetic systems, the distribution of the genetic distances between each population from cluster B and the Ju/'hoãnsi is significantly lower than those of populations in clusters A and C versus the Ju/'hoãnsi (mtDNA: $W = 94$, $P < 0.001$ and $W = 149$, $P < 0.001$; Y chromosome: $W = 128$, $P < 0.01$). Cluster B included 8 of the 14

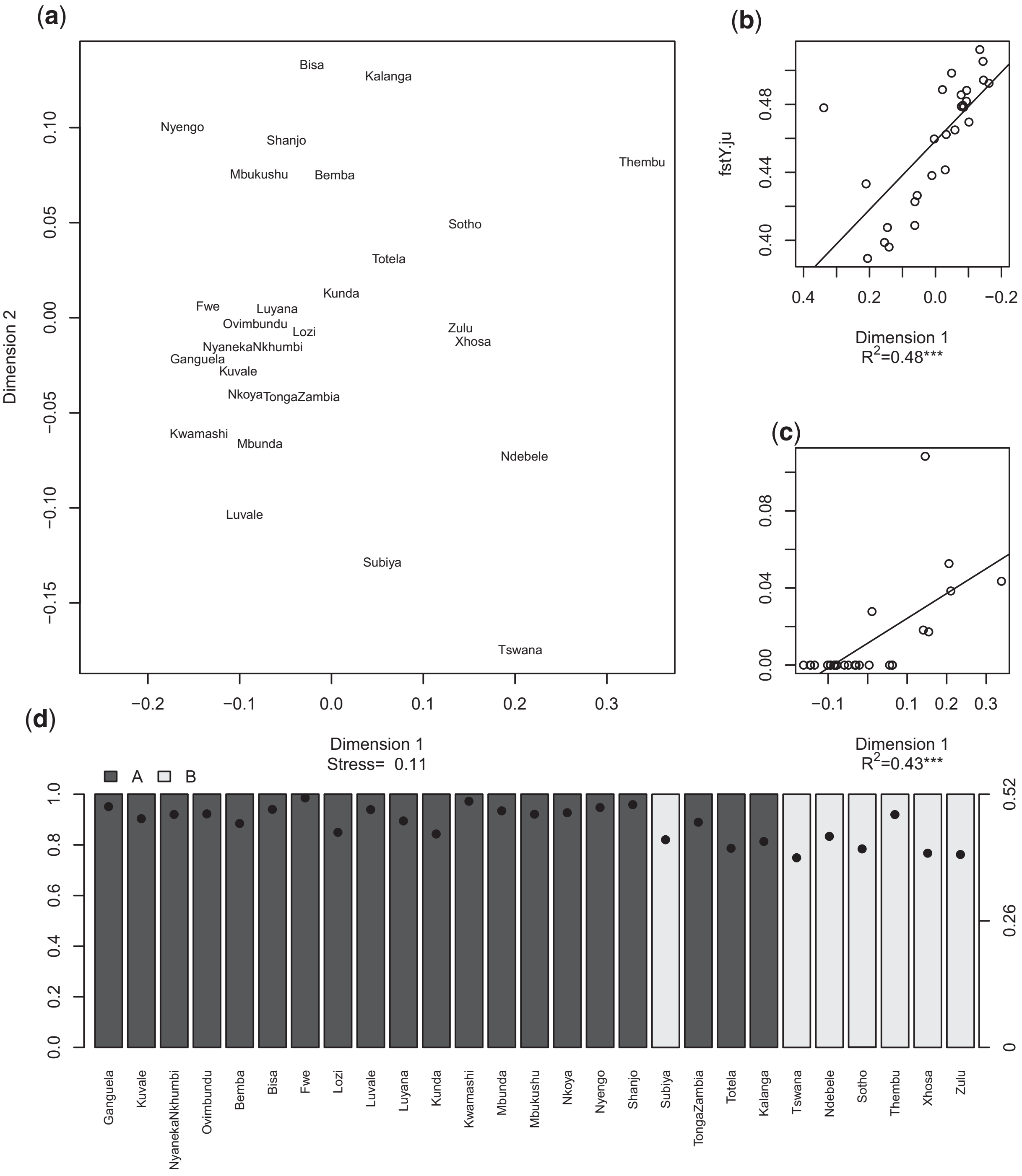

Fig. 3. (*a*) MDS plot based on the Reynolds' distance matrix for Y chromosome. (*b*, *c*) Correlation between first component of the MDS plot in (*a*) versus (*b*) the Reynolds' distance between each populations and the Ju/'hoãnsi and (*c*) the amounts of Y chromosome haplogroup A in the analyzed populations. (*d*) Cluster assignation probability as defined by DAPC analysis. Dots within each bar indicate the genetic distance of the indicated populations with the Ju/'hoãnsi.

populations showing the most extreme positioning along Dimension 2 in the MDS plot for mtDNA and the most extreme along axis 1 in the Y chromosome MDS plot (figs. 2*a* and 3*a*).

Migration Rates

In order to characterize the dynamics that have shaped the observed genetic variation further, we implemented a simulation approach that explored the frequency of haplotypes commonly present in Southern African foraging communities (here defined as HGH: Hunter-Gatherer-related haplogroups) detected in Bantu-speaking groups as a result of a given migration rate *m* (see Materials and Methods). For each population, we reported the interval of migration rates compatible with the amount of the HGH types observed (supplementary tables S6 and S7, Supplementary Material online, and fig. 4). When mtDNA data were considered, populations from the SE African region within Area 2 (fig. 1) were characterized by the largest estimate of migration rates. The Kuvale from Angola, the Ronga and the Ndau from Mozambique, and the Fwe from Zambia displayed the second largest set of estimates.

By focusing on the modal values for each distribution of estimated migration rates, populations from cluster B in the DAPC analysis show significantly higher migration rates than populations belonging to clusters A and C (Wilcoxon test: $W = 3$, $P < 0.001$ and $W = 0$, $P < 0.001$, respectively). Similarly significant differences were obtained when populations were grouped according to their location in relation to the ecoclimatic barrier characterizing the southernmost part of the African continent (fig. 1): Populations within Area 2 showed migration rate estimates significantly larger than populations from Area 1 (figs. 1 and 4 and supplementary table S6, Supplementary Material online; Wilcoxon test: $W = 0$, $P < 0.001$).

Such a clear distinction in migration rates among areas is not strongly present for the Y chromosome. In fact, the comparisons are not significant either using DAPC or geographic clustering criteria ($W = 63.5$, $P = 0.368$ and $W = 71$, $P = 0.5341$,

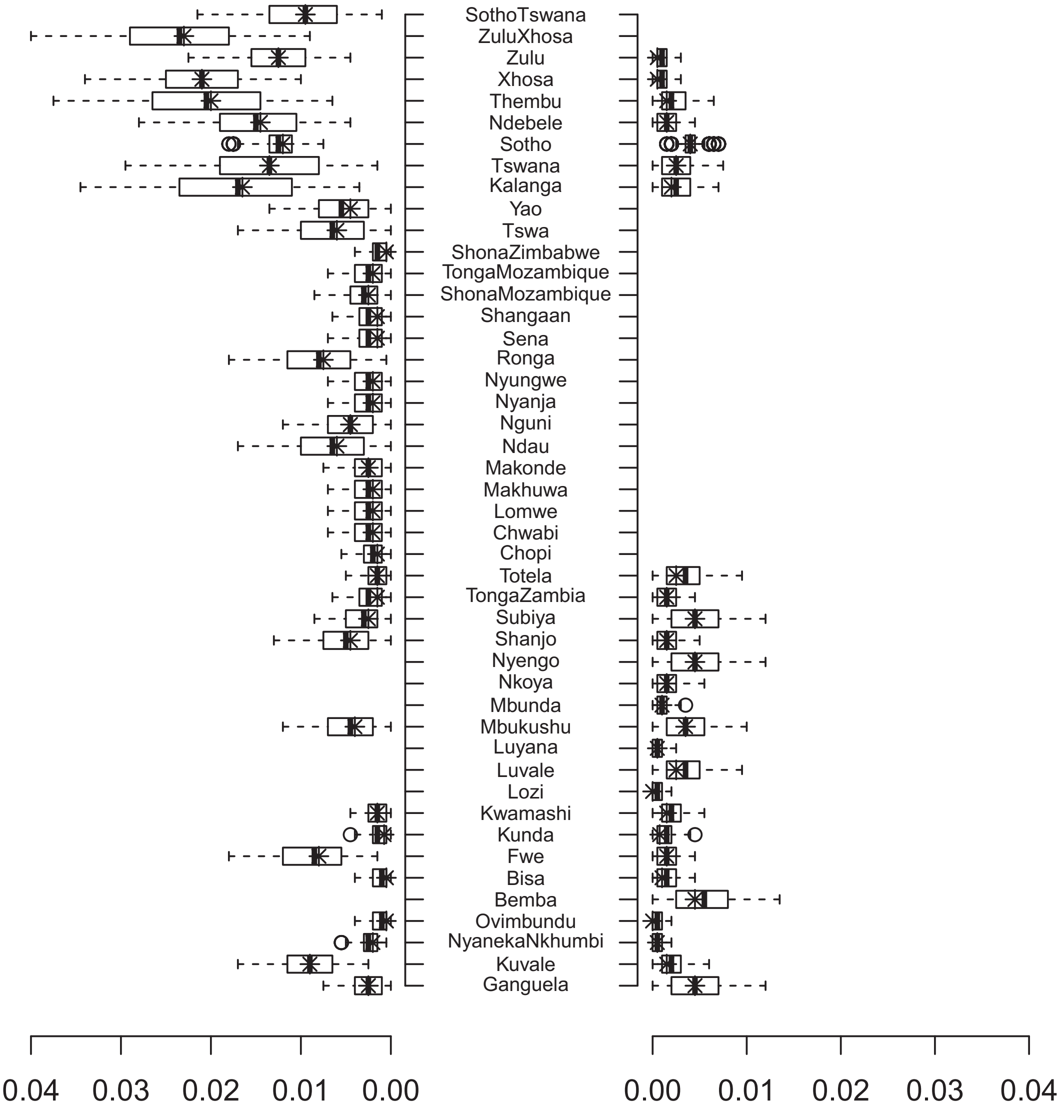

Fig. 4. Density of the migration rates for mtDNA (left) and Y chromosome (right) between a putative Khoesan population and the analyzed populations, based on simulation data (see main text). Boxes and whiskers represent the interquartile range and 1.5 times the interquartile range, respectively (circles refer to points that are outside this interval), while asterisks represent the modal value.

respectively; fig. 4 and supplementary table S7, Supplementary Material online), or they become marginally significant when the entire haplogroup B is taken into consideration ($W = 30.5$, $P = 0.015$ and $W = 34.5$, $P = 0.025$; data not shown). Interestingly, a higher migration rate for mtDNA than Y chromosome is observed when a paired Wilcoxon test is performed for the 20 populations analyzed for both the genetic systems ($V = 134$, $P < 0.001$). The comparison also remains significant when the migration rate estimates include the amount of Y chromosome haplogroup B present in the populations ($W = 210$, $P < 0.001$; data not shown).

Novel mtDNA and Y Chromosome Haplotypes in Southeastern African Bantu-Speaking Populations

We explored haplotypes identified in Lesotho and South Africa to investigate previously unreported lineages. The total number of novel mtDNA haplotypes was 81 (98 individuals), while 72 haplotypes (94 individuals) were found for the Y chromosome when compared across an extended regional database (see Materials and Methods; supplementary table S4, Supplementary Material online). Of these, a total of 40 haplotypes (51 individuals) belonged to mtDNA haplogroup L0d (supplementary table S4*a*, Supplementary Material online). Similarly, 10 A3b1 Y chromosome haplotypes (13 individuals) did not find a match (supplementary table S4*b*, Supplementary Material online). Interestingly, 56% of the previously unreported mtDNA L0d haplotypes showed a HaploGrep calling confidence below 90, even when the hypervariable segment II (HVR-II) region was included in the analysis (supplementary table S5, Supplementary Material online), probably due to the low number of sequences belonging to L0d present in the PhyloTree database.

The evolutionary relationships of haplotypes were investigated by constructing Median-Joining Networks, as described in the Materials and Methods section. A high degree of sharing between South Africa, Lesotho, and other regions characterizes the L0d mtDNA network (supplementary fig. S5, Supplementary Material online). Many haplotypes are unique to Lesotho or South Africa. Haplotypes from these countries are often linked to each other. A major cluster of closely related haplotypes dominates the network, with a small subset of haplotypes loosely clustering along a long, side branch. This subcluster contains haplotypes from Lesotho, South Africa, Namibia, Zambia, Mozambique, and Botswana, with sharing among populations occurring only for two haplotypes (one including South Africa, Namibia, Zambia, and Botswana; the other present only in South Africa and Namibia). Some further degree of structuring might be present within the mtDNA network: However, the low resolution offered by HVR-I data might be hiding such structure if indeed present, as whole mtDNA genome analysis has previously suggested (Barbieri, Vicente, et al. 2013). Hg A Y chromosome network is mostly characterized by haplotypes unique to specific areas, only three haplotypes being shared across countries: Two limited to South Africa and Lesotho and a third one present in these two countries plus Botswana. This pattern confirms a structured distribution of haplogroup A in Southern Africa, as previously suggested (Batini et al. 2011; supplementary fig. S6, Supplementary Material online).

Discussion

The Frontier Model and the Genetic Diversity of Southern African Bantu-Speaking Populations

It is known, from historical and archeological evidence, that Bantu-speaking farmers moved southwards and ultimately across most of sub-Equatorial Africa, starting approximately 5,000 years ago (Newman 1995). This movement accelerated significantly about 2,000 years ago, reaching southern Africa in the early centuries AD. Bantu-speaking farmers who settled in this region brought with them the knowledge of iron and a mixed farming economy based on cultivation of sorghum, pearl millet, and finger millet and the keeping of cattle, sheep, and goats. Archeological evidence suggests that this advance was rapid in the SC region and the easternmost part of South Africa's summer rainfall zone, providing little opportunity for interaction with indigenous populations, as predicted by a moving frontier scenario (Alexander 1984; Hall 1987). When a static frontier emerged and interaction became inevitable, the assimilation of the remaining foraging communities into the high density farmer populations resulted in an overall negligible contribution. By the mid-to-late first millennium AD, the southern expansion of these Bantu-speaking farmers had become constrained by the climatic and geographical conditions (specifically a combination of changing rainfall and temperature regimes defining the boundaries of the SE area; Area 2 in fig. 1), and thus they were unable to settle any further inland than the edge of the foothills of the Maloti/Drakensberg Mountains (Maggs 1976; Mitchell 2002; Mitchell et al. 2008), resulting in the development of a static frontier (Maggs 1980). Hunter-gatherer populations were still present at this time across large areas of Southern Africa, and archeological evidence has shown that much interaction seems to have occurred between them and the farmers. In KwaZulu-Natal, in particular, stone tools, bone arrow points, and ostrich eggshell beads of probable hunter-gatherer origin have been found at several early farming villages (Hall 1987; Mazel 1989), while farmer-associated ceramics, livestock, and iron have been found in some hunter-gatherer-associated rock shelter contexts (Carter and Vogel 1974; Mazel 1986; Mitchell et al. 2008). Hunter-gatherers were not necessarily subservient to farmers, and close, mutually beneficial relationships may have developed based on exchange, intermarriage, and perhaps also the supply of labor to farmers (Mazel 1989).

This static frontier lasted for approximately 1,000 years (Alexander 1984). Across this frontier, a significant degree of intermarriage and interaction nevertheless took place. One signal of this is the fact that several of the languages spoken in Lesotho and South Africa today include many click sounds of unquestionably Khoesan derivation (Herbert 1990; Güldemann and Stoneking 2008). These languages—Sesotho and the Nguni cluster of isiXhosa, isiZulu, isiNdebele, and siSwati—are totally atypical in this respect of the broader southeast Bantu group of languages spoken elsewhere in Botswana, Mozambique, South Africa, and Zimbabwe (Holden 2002; Lewis 2009). Specific cultural practices associated with divination in both Nguni- (Hammond-Tooke 1998, 1999) and Sesotho-speaking (Tesele 1994) populations also have a clear San source. Moreover, in recent centuries, at least one Xhosa-speaking group (the Mpondomise) employed San rainmakers and other Bantu speaker-associated items of material culture featured in San rock art (Loubser and Laurens 1994; Challis 2012). Once livestock assumed a more important role in the economy and crossing the ecoclimatic barrier became feasible by a switch in the crops used, these admixed communities were able to colonize new areas and expand further within the SE African region (Area 2 in fig. 1).

This scenario goes some way toward explaining our genetic findings from Lesotho and South Africa. The correlation between genetics and geographic distances support the conclusions reported by others (Tishkoff et al. 2007; de Filippo et al. 2012) regarding the role played by the demic dispersal of Bantu-speaking populations in shaping the distribution of genetic variation in sub-Saharan Africa. However, our results suggest that in addition to this dispersal, other processes have influenced the diversity observed in Bantu-speaking groups from Southern Africa (fig. 1). Despite a broad correspondence between the MDS plot and geographic provenance, a group of populations (including all those from Area 2 plus a few others from Zambia, Angola, and Mozambique) tends to separate along the second axis of the MDS according to their higher affinity (reflected by a lower genetic distance), with the Ju/'hoãnsi in Namibia and the frequency of mtDNA haplogroups commonly present in Southern African foraging groups, L0d and L0k (fig. 2*a*). The influence of these mtDNA haplogroups on the association between these populations is further supported by correspondence analysis (supplementary fig. S1, Supplementary Material online). The distinct genetic pattern of populations from Area 2 is further suggested by them forming one of the three groups identified by the DAPC analysis, (fig. 2*d*) and by their Reynolds' distance values versus the Ju/'hoãnsi being significantly smaller than those of populations belonging to the other two groups (fig. 2*d*). We interpreted these results as supporting a stronger signature of admixture with hunter-gatherers in populations from this region (SE) than those present in other areas, namely the SC region, as also supported by migration rate estimates based on simulation data (fig. 4). This interpretation is consistent with the predictions of the frontier model. The persistence of foraging communities during the static phase of the Bantu dispersal along the SE region (caused by the ecoclimatic barrier discussed above; fig. 1) facilitated the interaction and assimilation of foragers into farming communities. The signature of this process has been preserved and highlighted by the subsequent population expansion made possible by the colonization of areas across the barrier (moving frontier phase). The lack of similarly strong signals in other areas (although local differences can be present within regions, see below) could be the result of the marginalization of foragers by Bantu-speaking farmers during the rapid colonization characterizing the moving frontier phase in much of Southern Africa. Relevant interactions with remaining foraging groups occurred only later on, when the occupation of the available space was completed and a static frontier developed, the evidence of possible assimilation episodes being diluted by the high density, and large population size reached by farming communities at that point in time.

Our analysis assumed that the genetic makeup of extant Southern African foraging populations (or those that have been so until recently) can be used to infer the composition of the pre-agricultural communities once present all across Southern Africa. Haplotypes commonly found in Khoesan-speaking populations have been found across the region in Bantu-speaking groups whenever archeological, linguistic, and ethnographic data supported a scenario of contact and exchange between foragers and agriculturalists. It has been suggested, for example, that the recent identification of novel mtDNA branches within L0k haplogroups in Bantu-speaking populations from Zambia might represent relics from groups previously living in these regions (Barbieri, Vicente, et al. 2013). Our findings of previously unreported, geographically localized L0d and A3b1 haplotypes in Bantu-speaking populations provide further evidence of the widespread distribution of these types and support the existence of some degree of population structure among the pre-Bantu groups originally inhabiting the area, as expected due to drift and isolation of groups geographically separated by thousands of kilometers who had been living in the region for tens of thousands of years (Barbieri, Vicente, et al. 2013; Barbieri et al. 2014; Pickrell et al. 2014). Ancient population structure is also supported by the identification of A3b1 but not A2 Y chromosome haplotypes in Bantu-speaking groups (despite their high prevalence in extant Southern African foraging communities; Underhill et al. 2000; Wood et al. 2005; Batini et al. 2011; this work) and the differential distribution of L0d/L0k mtDNA types in these areas, as previously suggested (Batini et al. 2011; Barbieri, Vicente, et al. 2013). Further analysis of these novel haplotypes could provide insights into the genetics of the indigenous occupiers of the region, in particular the whole mtDNA genomes of the haplotypes showing novel mutation combinations as already reported for other regions and subhaplogroups (Barbieri, Vicente, et al. 2013b). In addition, the forthcoming combination of genome data and novel biostatistical approaches will contribute to the characterization of these admixture events in terms of their dates, the admixing sources, and their relative contributions in a deeper detail, as already demonstrated in previous investigations (Hellenthal et al. 2014; Pickrell et al. 2014).

We note that the interpretation of the pattern of genetic variation among Bantu-speaking populations is often further complicated by the presence of multiple layers of farmer occupancy in the same areas, something that potentially occurred in South Africa and Lesotho. Here, the pottery made by the Sotho/Tswana- and Nguni-speaking populations that dominate this area today cannot be derived from the ceramics associated with first millennium AD farmers, and several other changes in settlement pattern and ritual practice are also evident (Maggs 1994; Mitchell 2002; Huffman 2007).

Instead, the earliest Nguni-associated pottery (Blackburn, which appears in KwaZulu-Natal around AD 1000) and the first Sotho/Tswana-linked ceramics (Moloko, which appears in Limpopo and Mpumalanga about AD 1300) mark a major disjunction in the archeological record interpreted as the arrival of people speaking these languages from further north in East Africa (Huffman 1988). Linguistic data support this scenario by indicating that first millennium AD farmers in KwaZulu-Natal likely spoke a Shona-like language, traces of which survive in modern Nguni (Ownby 1988). Ancestral Shona speakers may also have moved north into the Limpopo Valley and then beyond this into Zimbabwe from around AD 1000 (Huffman 2007).

The relationships between newly arriving ancestral Nguni and Sotho/Tswana speakers and the farming communities that were already present in southernmost Africa remain a topic of active research in archeology. How the change in the makeup of the region's agropastoralist population affected relations with surviving hunter-gatherer groups also requires further study, but the restriction of Khoesan-derived click sounds to the Nguni languages (especially isiXhosa) and South Sotho implies a different intensity of interaction than between hunter-gatherers and the ancestors of the Shona, Venda, and more northerly Sotho/Tswana-speaking communities (Bostoen and Sands 2012). The distinctive features of divination practices among modern Nguni speakers also reflect sustained and intimate interaction with hunter-gatherers, likely mediated by hypergamic marriage (Hammond-Tooke 1998, 1999). It follows from this that both the presence and the temporal duration of the static frontier were critical in shaping the degree of interaction and assimilation of hunter-gatherers into farming communities.

Finally, we note that other populations (the Kuvale from Angola, the Fwe from Zambia, and the Ronga and Ndau from Mozambique) display increased amounts of L0d/L0k and/or closer affinity with the Ju/'hoãnsi and higher migration rates with foragers. It is not unexpected that the interactions between farmers and foraging communities occurred differently in various regions as a result of local dynamics and population-specific demographic histories, as already suggested for some of these groups (Coelho et al. 2009; Barbieri, Butthof, et al. 2013; Barbieri, Vicente, et al. 2013; Barbieri et al. 2014). Our results, while providing a novel interpretative framework for the genetic variation displayed by populations from Lesotho and South Africa, do not necessarily imply that the frontier model is the only explanation for the occurrence of gene flow between farmers and hunter-gatherers.

Gender-Biased Gene Flow and the Role of Women in Cross-Cultural Admixture

The evidence of gene flow between farmers and foraging communities reported here exhibits similarities with previous investigations: The characterization of admixture as a nonrandom, gender-biased process. In addition to different amount of HG lineages in Bantu-speaking populations from different areas, we in fact found evidence for differences in frequencies of Y chromosome and mtDNA from foragers. The female migration rate from foraging into Bantu communities is considerably higher than that of men (fig. 4 and supplementary tables S6 and S7, Supplementary Material online). This trend toward hypergamy has previously been reported in other African populations (Cavalli-Sforza 1986; Destro-Bisol et al. 2004; Destro-Bisol 2005; Wood et al. 2005; Quintana-Murci et al. 2008), suggesting that similar sociocultural dynamics might have also existed in southernmost Africa, favoring the assimilation of female more than male hunter-gatherers into Bantu-speaking farming communities, as suggested by previous investigations (Coelho et al. 2009; Quintana-Murci et al. 2010; Barbieri, Butthof, et al. 2013; Schlebusch et al. 2013). The tendency for intermarriage to favor hunter-gatherer women over men can be explained in terms of a prevailing southern Bantu emphasis on patrilineal descent and the need to pay bridewealth for wives; hunter-gatherers will not have had access to the livestock needed for this.

A Revised Model for the "Bantu Expansion"

The relative linguistic homogeneity observed across most of Africa south of the Equator has been the subject of much speculation in an attempt to generate a model to explain the widespread distribution of the Bantu languages. Together with archeology, population genetics has provided complementary evidence contributing to the characterization of the demographic dynamics that have been associated with such linguistic dispersal. Data from a variety of genetic systems, including Y chromosome and mtDNA haplogroup distribution, and variation of autosomal SNPs and STRs, have indicated a broad genetic homogeneity across Bantu-speaking populations, providing support for a demic model for the dispersal of the Bantu languages (Tishkoff et al. 2009; de Filippo et al. 2012). More recently, attention has been given to the fine-scale characterization of this process; work by de Filippo et al. (2012) represent one of the latest attempts in this direction. By combining extensive genetic and linguistic data sets, these authors have proposed a model indicating a later split between populations speaking Eastern and Western Bantu languages (the so-called "late split model"). Within this scenario, the close affinity observed across Bantu-speaking groups suggested minimal language shift and/or gene flow from groups previously resident in the areas where the expansion took place. Our results broadly fit this demic scenario. However, the evidence we reported for more extensive gene exchange in certain areas can be reconciled with de Filippo et al. by the fact that these authors almost completely lack populations from Area 2 (fig. 1), making it impossible for them to detect the signal which we have reported here. Their results also contrasted with those reported by Sikora et al. (2011), who suggested a very significant non-Bantu component in a sample of Mozambicans analyzed using genome-wide SNP data. Our results cannot reject the hypothesis that some gene flow between farmers and foragers might have occurred in the SC region (fig. 4). However, the degree of gene exchange was probably much lower than that implied by the STRUCTURE analysis of Sikora et al. Some differentiation among Bantu-speaking groups is expected as a result of the dispersal process independently of the degree of gene flow from non-Bantu-speaking populations, as also suggested by the significant Mantel test and the results of the DAPC analysis (fig. 2*d*) differentiating the Angola and Zambian populations belonging to Bantu major zones R, M, K (Guthrie 1971) from the other Southern African populations reported here (major zones N, S, P). In summary, the evidence for gene flow presented here for SE Southern Africa complements and refines the no-admixture model put forward by de Filippo et al. (2012). The analysis presented here also questions the proposal by Sikora et al. of a large non-Bantu component in Bantu-speaking populations from Mozambique. Our results suggest a more complex dispersal model and point to locally defined, socioecological-driven episodes of gene flow during the so-called Bantu expansion. Future investigations aimed at developing a comprehensive framework for the demographic modalities of the dispersal of Bantu-speaking populations should incorporate the frontier model discussed here and properly consider the admixture dynamics that this implies.

Materials and Methods

Samples

Lesotho Samples

Saliva samples of unrelated men from Lesotho were collected in October and November 2009 using Oragene DNA sample collection kits (DNA Genotek, Inc., Ottawa, ON), and extracted using the manufacturer's protocols. Participants were healthy adults, and appropriate informed consent was obtained. The research was approved by the Oxford Tropical Research Ethics Committee (OxTREC), the Lesotho Ministry of Health and Social Welfare, the Lesotho Ministry of Local Government, and the Lesotho Ministry of Tourism, Environment and Culture.

Three hundred forty-four samples were collected from six different locations within the country, including three locations from the Lesotho lowlands and three from the highlands (Marks et al. 2012; fig. 1). Ethnic and linguistic information was also collected about the donors, their parents, and their grandparents, where known. The total sample included individuals from various ethno-linguistic groupings, the majority of whom were Basotho, Ndebele, and Thembu. For analytical purposes, the ethnic information of the maternal grandmother was used for mtDNA analyses, and the ethnic information of the paternal grandfather for the Y chromosome analyses.

South Africa Samples

DNA samples from 58 Xhosa (Leat et al. 2004, 2007) and 54 Zulu individuals sampled in Cape Town were collected by the Forensic DNA Laboratory, Department of Biotechnology, University of the Western Cape in Cape Town, South Africa.

Namibia Samples

For comparison, we also included 58 DNA samples collected from the Ju/'hoãnsi San in the Tsumkwe area in 2010 using Oragene DNA sample collection kits (ethical approval by OxTREC and the Namibian Ministry of Health and Social Services).

Genotyping

The mtDNA HVR-I was amplified using primers L15997 (5′-CACCATTAGCACCCAAAGCT-3′), H00017 (5′-CCCGTGAGTGGTTAATAGGGT-3′), and H16401 (5′-TGATTTCACGGAGGATGGTG-3′) (Pereira et al. 2001), and variable positions were determined within positions 16040–16519 (supplementary table S3*a*, Supplementary Material online) for 510 individuals (Sotho = 287, Ndebele = 33, Thembu = 24, Xhosa = 54, Zulu = 54, Ju/'hoãnsi = 58). Sequencing was performed using BigDye Terminator Chemistry version 1.1 (Applied Biosystems) according to the manufacturer's protocol, and was followed by Shrimp Alkaline Phosphatase (SAP) and ethanol–sodium acetate purification. Purified products were sequenced on an ABI 3730x1 Genetic Analyser (Applied Biosystems). Each sequence was assigned to a haplogroup using information from previous studies (Behar et al. 2008; van Oven and Kayser 2009) (supplementary tables S2*a* and S3*a*, Supplementary Material online). Haplogroup assignment was verified using the HaploGrep Web site (based on PhyloTree version 15; van Oven and Kayser 2009; Kloss-Brandstätter et al. 2011). One hundred four sequences provided HaploGrep scores below 90. Sequences were confirmed by rechecking the related electropherograms and resequencing. HaploGrep-reported problems with these samples were due to the inability of assigning the sequences to sublineages. In all the cases, these haplotypes had in fact the mutations necessary to assign them to a specific haplogroup. The HaploGrep results are probably due to the fact that the sequences exhibited additional new mutations/combination of variants that have not been previously reported, meaning that the program could not successfully match them with known sequences. For example, a number of samples were defined by HaploGrep as L0d1'2, as they had the necessary mutations for L0d (16129, 16187, 16189, 16223, 16230, 16243, 16311), and also a reversion at position 16278 to place them in the L0d1'2 branch, but not the other required polymorphisms to distinguish between either L0d1 or L0d2. Additionally, these samples had a number of extra mutations that the program calls "local private mutations" which prevented matches with the database. Therefore for simplicity, all samples indicated as belonging to one of the L0d branches were included as L0d* in analyses. mtDNA haplogroup frequencies are reported in supplementary table S2*a*, Supplementary Material online. Haplotypes and their haplogroup assignation as suggested by HaploGrep are reported in supplementary table S3*a*, Supplementary Material online. Of the total number of haplotypes identified, 37.3% reported a HaploGrep score below 90. These haplotypes were equally divided among those belonging to L0d (28) and other haplogroups (30), showing similar frequencies in their relative groups (37.7% and 36.5%). The lowest scores were 66 and 74 in the two groups, respectively. Sequences found to be unique in SC Bantu-speaking populations belonging to L0d and L0k haplogroups (12 and none, respectively) and having a HaploGrep score below 90 were extended to include the HVR-II. Amplification and sequencing were carried out as for HVR-I, but using forward primer 16555L (5′-CCCACACG

TTCCCCTTAAAT-3′) and reverse primer 599H (5′-TTGAGGA GGTAAGCTACATA-3′) for the initial amplification reaction, and additionally primer 408H (5′-CTGTTAAAAGTGCATACC GCCA-3′) for sequencing (Cerezo et al. 2009). Haplotypes combining both the HVR-I and HVR-II mutations were checked again using HaploGrep.

Y chromosome haplotype variation was explored by genotyping 17 STRs using the Yfiler multiplex kit (Mulero et al. 2006) for 411 individuals (Sotho = 179, Ndebele = 48, Thembu = 21, Xhosa = 57, Zulu = 51, Ju/'hoãnsi = 53). Alleles were called by the inclusion of an internal size standard and comparison with a reference allelic ladder as implemented in the GeneMapper software (Applied Biosystems) (supplementary table S3*b*, Supplementary Material online). Haplogroup analysis was conducted by genotyping 12 SNP markers (P97, P247, M150, M112, P248, M181, M51, M96, M206, M89, M267, M304; supplementary fig. S7, Supplementary Material online) in one multiplex reaction combining previously published protocols (Onofri et al. 2006; Batini et al. 2011) for 511 individuals (Sotho = 267, Ndebele = 52, Thembu = 23, Xhosa = 57, Zulu = 54, Ju/'hoãnsi = 58). Haplogroup frequencies are reported in supplementary table S2*b*, Supplementary Material online. Nomenclature used for NRY SNP markers discussed in this work follows the Y Chromosome Consortium (2002) indication (Karafet et al. 2008).

Within- and between-Population Diversity

Previous analyses of the samples collected from Lesotho showed no evidence of population structure among the sampled areas (Marks et al. 2012). For this reason, the Sotho samples were considered as representing a single homogenous population and analyzed accordingly.

Within-population summary statistics were calculated using Arlequin 3.5 software (Excoffier and Lischer 2010; table 1). For Y chromosome, all the within-population analyses were performed on 15 loci haplotypes after the removal of DYS385 marker.

To provide a comprehensive picture of the genetic landscape in the SC and SE regions of the African continent, we collected data from the literature on Angola, Zambia, Zimbabwe, Mozambique, Botswana, and South Africa, creating a data set of 1,473 (from 39 populations) and 1,189 individuals (from 27 populations) for mitochondrial and Y chromosome, respectively (fig. 1 and supplementary table S1, Supplementary Material online). For the latter data set, only 10 loci were considered (DYS19, DYS389I, DYS389II—alleles reported after subtracting the number of repeats estimated for DYS389I—DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439).

Great circle geographic distance matrix was obtained by the "rdist.earth" function implemented in fields R package (Nychka et al. 2013) using approximate sampling locations (fig. 1).

Mantel test (Mantel 1967; Mantel and Valand 1970) between geographic and genetic distances (Reynolds et al. 1983) was performed using the R package vegan (Oksanen et al. 2013), performing 10,000 permutations. Population relationships were explored using multidimensional scaling analysis based on haplotypic data using the isoMDS function from MASS R package, based on the Reynolds' (Reynolds et al. 1983) distance estimated using the ade4 R package (Thioulouse et al. 1997). The significance of the stress value was evaluated as described in Sturrock and Rocha (2000).

Correspondence analysis was carried out with CA R package (Nenadic and Greenacre 2007) using haplogroup frequencies.

Regression Analysis

Hunter-gatherer-related haplogroups (HGH) present in Southern Africa before the arrival of Bantu-speaking groups are generally identified as belonging to haplogroups L0d/L0k for mtDNA and A/B2b for the Y chromosome (Chen et al. 2000; Pereira et al. 2001, 2002; Salas et al. 2002; Behar et al. 2008; Coelho et al. 2009; de Filippo et al. 2010, 2011; Batini et al. 2011; Barbieri, Vicente, et al. 2013). Unfortunately, only a subset of the samples were characterized for the B2b defining SNPs, making impossible the distinction between this haplogroup and others whose distribution might be affected by other demographic processes (Batini et al. 2011). For this reason, we focus on haplogroup A but also reported results when all haplogroup B haplotypes were considered as HGH.

We tested for significant correlation between 1) dimensions of the MDS analysis described above and the genetic distance (Reynolds et al. 1983) of each population against Ju/'hoãnsi and 2) the HGH frequency and the first two dimensions of the multidimensional scaling analysis (figs 2 and 3).

Discriminant Analysis of Principal Components

In order to investigate the genetic structure among the analyzed populations, we performed DAPC (Jombart et al. 2010). This method consists in the discriminant analysis of the data after their transformation by PC analysis, taking into account a priori affiliation for each element. This approach allows the analysis of uncorrelated variables maintaining most of the genetic information (Jombart et al. 2010).

We retained the principal components explaining at least 80% of the total variance, using the R software adegenet package (Jombart et al. 2009, 2010). The best number and composition of the clusters were inferred using the BIC analysis, as implemented in the find.clusters function retaining all the principal components (supplementary fig. S1, Supplementary Material online). The BIC distribution was characterized for both mtDNA and Y chromosome data by an initial decrease for low numbers of clusters, an increase for medium numbers of clusters, and a final decay. Because our target was to summarize the genetic structure displayed by our populations, we retained as the best number of clusters the value associated with the smallest BIC value before the final decay. We then plotted the first discriminant functions of each analysis and generated a composition plot where the probability of assignment to each cluster is reported (supplementary fig. S2, Supplementary Material online). Analysis was based on the frequencies of mutations of the analyzed genetic systems. In detail, we used HVR-I SNP frequencies for mtDNA sequences, and STR allelic frequencies for Y chromosome haplotypes. Alleles at different loci are not randomly distributed as they are linked within haplotypes. Given the association between haplotypes and haplogroups, this approach indirectly allows for the recovery of information of population haplogroup composition without relying on direct haplogroup assignation. Such an approach has been already applied (Montano et al. 2013).

Simulations

As we were interested in detecting common patterns of migration more than fine characterizing the absolute intensity of the gene flow, we implemented admixture simulations described in Barbieri, Butthof, et al. (2013). We considered two populations ("San" and "Bantu") of constant size 10,000 individuals, with the assumption that the two were characterized by completely different sets of haplotypes. The Bantu population was allowed to receive migrants from the San population at a rate m for $t = 29$ generations (the approximate time of the Bantu-speaking groups arrival in Southern Africa; Mitchell 2002), considering an average generation time of 30 years (Fenner 2005). After t generations, the frequency of the San types in the admixed populations was calculated and recorded. This process was repeated 1,000 times for migration rates m sampled in the range 0.00–0.07 with incremental increases of 0.0005 (total number of simulations = 1,410,000). The amount of Khoesan haplotypes present in the SE and SC African samples was used to predict the range of associated migration rate m by referring to the simulation data (see Results).

Novel Y Chromosome and mtDNA Haplotypes

mtDNA and Y chromosome haplotypes of novel samples of Bantu-speaking populations from Lesotho and South Africa presented in this work were screened for matches to a database assembled from a variety of Southern African populations (Salas et al. 2002; Alves et al. 2003; Kido et al. 2006; Castrì et al. 2009; Coelho et al. 2009; de Filippo et al. 2010, 2011; Melo et al. 2010; Quintana-Murci et al. 2010; Batini et al. 2011; Schlebusch et al. 2011, 2013; Barbieri, Butthof, et al. 2013; Barbieri, Vicente, et al. 2013; including our newly reported set of Ju/'hoãnsi from Namibia). In order to maximize the coverage of the regional variation, mtDNA comparisons were restricted to the 16040–16383 range of the HVR-I and to 10 STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439) for the Y chromosome. A total of 801 haplotypes (2,478 individuals) from 89 populations and 966 haplotypes (1,596 individuals) from 46 populations were presented in the mtDNA HVR-I and Y chromosome databases, respectively.

Median-Joining Networks of haplogroup L0d (HVR-I range: 16,040–16,383 bp; supplementary fig. S5, Supplementary Material online) and haplogroup A (10 locus STR haplotypes; supplementary fig. S6, Supplementary Material online) were constructed according to Zalloua et al. (2008) using the program NETWORK version 4.6.1.2 (www.fluxus-engineering.com, last accessed December 2013). When constructing networks, the default value (10) was given to each HVR-I site, while each STR locus was weighted according to its variance; the weight of the ith STR was calculated as $10 \times V_m/V_i$, where V_m is the mean variance of all STRs and V_i is the variance of the ith STR (Bosch et al. 2006).

Supplementary Material

Supplementary figures S1–S7 and tables S1–S7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank N. Matlanyane and N. Qokolo for their contribution to the collection of the Lesotho samples; Simon Marks for help with the collection of the Namibian samples; and the Lesotho and Namibian Ministries of Health and Social Welfare, Local Government, Tourism, Environment and Culture for their support. We would also like to thank all the people who made this work possible by contributing with their DNA samples and all the individuals who facilitated the organization of the field work and the collection of the samples. This project was funded by the Boise Fund and the Wenner-Gren Foundation. S.J.M. was a recipient of a Boise Fund Studentship and H.L. was funded by a Rhodes Scholarship.

References

Alexander JA. 1977. Hunters: gatherers and first farmers beyond Europe: an archaeological survey. Leicester (United Kingdom): Leicester University Press.

Alexander JA. 1984. Early frontiers in southern Africa. In: Hall M, Avery G, Avery DM, Wilson ML, Humphreys AJB, editors. Frontiers: Southern African archaeology today. Oxford: B.A.R. International Series. p. 12–23.

Alves C, Gusmão L, Barbosa J, Amorim A. 2003. Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. *Forensic Sci Int.* 134:126–133.

Ammerman A, Cavalli-Sforza L. 1973. The explanation of culture change: models in prehistory: proceedings. Pittsburgh (PA): University of Pittsburgh Press.

Barbieri C, Butthof A, Bostoen K, Pakendorf B. 2013. Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *Eur J Hum Genet.* 21:430–436.

Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Stoneking M, Pakendorf B. 2014. Unraveling the complex maternal history of Southern African Khoisan populations. *Am J Phys Anthropol.* 153:435–448.

Barbieri C, Vicente M, Rocha J, Mpoloka SW, Stoneking M, Pakendorf B. 2013. Ancient substructure in early mtDNA lineages of southern Africa. *Am J Hum Genet.* 92:285–292.

Batini C, Ferri G, Destro-Bisol G, et al. 2011. Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Mol Biol Evol.* 28: 2603–2613.

Behar DM, Villems R, Soodyall H, et al. 2008. The dawn of human matrilineal diversity. *Am J Hum Genet.* 82:1130–1140.

Bohannan P, Plog F. 1967. Beyond the frontier: social process and cultural change. Garden City (NY): Published for the American Museum of Natural History [by] Natural History Press.

Bosch E, Calafell F, González-Neira A, et al. 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet.* 70:459–487.

Bostoen K, Sands B. 2012. Clicks in south-western Bantu languages: contact-induced vs. language-internal lexical change. In: Brenzinger M, editor. In: Proceedings of the 6th World Congress of African Linguistics Cologne 2009. Köln (Germany): Rüdiger Köppe Verlag. p. 129–140.

Burt AL. 1940. The frontier in the history of New France. *Rep Annu Meet.* 19:93–99.

Carter PL, Vogel JC. 1974. The dating of industrial assemblages from stratified sites in Eastern Lesotho. *Man* 9:557–570.

Castrì L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, Paoli G, Pettener D, Luiselli D. 2009. mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa. *Am J Phys Anthropol.* 140:302–311.

Cavalli-Sforza LL. 1986. African pygmies. New York: Academic Press.

Cerezo M, Bandelt HJ, Martín-Guerrero I, Ardanaz M, Vega A, Carracedo A, García-Orad A, Salas A. 2009. High mitochondrial DNA stability in B-cell chronic lymphocytic leukemia. *PLoS One* 4:e7902.

Challis S. 2012. Creolisation on the nineteenth-century frontiers of Southern Africa: a case study of the AmaTola "Bushmen" in the Maloti-Drakensberg. *J S Afr Stud.* 38:265–280.

Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC. 2000. mtDNA variation in the South African Kung and Khwe—and their genetic relationships to other African populations. *Am J Hum Genet.* 66:1362–1383.

Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol Biol.* 9:80.

Cruciani F, Santolamazza P, Shen P, et al. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet.* 70:1197–1214.

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62:1908–1920.

de Filippo C, Barbieri C, Whitten M, et al. 2011. Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol Biol Evol.* 28:1255–1269.

de Filippo C, Bostoen K, Stoneking M, Pakendorf B. 2012. Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc Biol Sci.* 279:3256–3263.

de Filippo C, Heyn P, Barham L, Stoneking M, Pakendorf B. 2010. Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. *Am J Phys Anthropol.* 141:382–394.

Destro-Bisol G. 2005. Genetic variation and social structure: a case-study from Africa. *Hum Evol.* 20:93–98.

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

Diamond JM. 1997. Guns, germs, and steel: the fates of human societies. New York: W.W. Norton & Co.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10:564–567.

Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol.* 128:415–423.

Gill SJ. 1993. A short history of Lesotho from the late stone age until the 1993 elections. Morija (Lesotho): Morija Museum & Archives.

Gomes V, Sánchez-Diz P, Amorim A, Carracedo A, Gusmão L. 2010. Digging deeper into East African human Y chromosome lineages. *Hum Genet.* 127:603–613.

Güldemann T, Stoneking M. 2008. A historical appraisal of clicks: a linguistic and genetic population perspective. *Ann Rev Anthropol.* 37:93–109.

Guthrie M. 1971. Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages. Bantu prehistory, inventory and indexes. Farnborough (United Kingdom): Gregg Press.

Hall TD. 1987. Native Americans and incorporation: patterns and problems. *Am Indian Cult Res J.* 11:1–30.

Hammond-Tooke WD. 1998. Selective borrowing? The possibility of San shamanistic influence on Southern Bantu divination and healing practices. *S Afr Archaeol Bull.* 53:9–15.

Hammond-Tooke WD. 1999. Divinatory animals: further evidence of San/Nguni borrowing? *S Afr Archaeol Bull.* 54:128–132.

Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A genetic atlas of human admixture history. *Science* 343: 747–751.

Herbert RK. 1990. The sociohistory of clicks in Southern Bantu. In: Mesthrie R, editor. Language in South Africa. Cambridge (MA): Cambridge University Press. p. 297–315.

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

Huffman TN. 1988. Ngwenani ya themeli. In: Evers TM, Huffman TN, Wadley L, editors. Guide to archaeological sites in the transvaal. Johannesburg (South Africa): University of the Witwatersrand. p. 52–55.

Huffman TN. 2007. Handbook to the iron age: the archaeology of pre-colonial farming societies in Southern Africa. Scottsville (South Africa): University of KwaZulu-Natal Press.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94.

Jombart T, Pontier D, Dufour AB. 2009. Genetic markers in the playground of multivariate analysis. *Heredity (Edinb)* 102:330–341.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.

Kido A, Fujitani N, Hara M, Kimura H. 2006. Genetic data of 16 Y-chromosomal short tandem repeat loci in Africans from South Africa. *J Forensic Sci.* 51:1414–1416.

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

Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat.* 32:25–32.

Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr Biol.* 13:464–473.

Leat N, Benjeddou M, Davison S. 2004. Nine-locus Y-chromosome STR profiling of Caucasian and Xhosa populations from Cape Town, South Africa. *Forensic Sci Int.* 144:73–75.

Leat N, Ehrenreich L, Benjeddou M, Cloete K, Davison S. 2007. Properties of novel and widely studied Y-STR loci in three South African populations. *Forensic Sci Int.* 168:154–161.

Lewis MP. 2009. Ethnologue: languages of the world. Dallas (TX): SIL International.

Loubser J, Laurens G. 1994. Depictions of domestic ungulates and shields: hunter-gatherers and agro-pastoralists in the Caledon Valley area. In: Dowson T, Lewis-Williams JD, editors. Contested images: diversity in southern African rock art research. Johannesburg: Wits University Press. p. 83–118.

Maggs T. 1976. Iron age communities of the southern Highveld. Pietermaritzburg (South Africa): Council of the Natal Museum.

Maggs T. 1980. The iron age sequence south of the Vaal and Pongola rivers: some historical implications. *J Afr Hist.* 21:1–15.

Maggs T. 1994. The early iron age in the extreme south: some patterns and problems. *Azania* 29/30:171–178.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.

Mantel N, Valand RS. 1970. A technique of nonparametric multivariate analysis. *Biometrics* 26:547–558.

Marks SJ, Levy H, Martinez-Cadenas C, Montinaro F, Capelli C. 2012. Migration distance rather than migration rate explains genetic diversity in human patrilocal groups. *Mol Ecol.* 21:4958–4969.

Mazel AD. 1986. Mbabane Shelter and eSinhlonhlweni Shelter: the last two thousand years of hunter-gatherer settlement in the central Thukela Basin, Natal, South Africa. *Ann Natal Museum.* 27:389–453.

Mazel AD. 1989. People making history: the last ten thousand years of hunter-gatherer communities in the Thukela Basin. *Natal Museum J Human.* 1:1–168.

Melo MM, Carvalho M, Lopes V, Anjos MJ, Serra A, Vieira DN, Sequeiros J, Corte-Real F. 2010. Genetic study of 15 STRs loci of Identifiler system in Angola population. *Forensic Sci Int Genet.* 4:e153–e157.

Mitchell P. 2002. The archaeology of Southern Africa. Cambridge (MA): Cambridge University Press.

Mitchell P. 2009. Hunter—gatherers and farmers: some implications of 2000 years of interaction in the Maloti—Drakensberg region of southern Africa. *Senri Ethnol Stud.* 73:15–46.

Mitchell PJ, Plug I, Balley GN, Woodborne S 2008. Bringing the Kalahari debate to the mountains: late first millennium AD hunter-gatherer/farmer interaction in highland Lesotho., Before Farming [Internet], 2008/2 article 4.

Montano V, Marcari V, Pavanello M, Anyaele O, Comas D, Destro-Bisol G, Batini C. 2013. The influence of habitats on female mobility in Central and Western Africa inferred from human mitochondrial variation. *BMC Evol Biol.* 13:24.

Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL, Hennessy LK. 2006. Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci.* 51:64–75.

Nenadic O, Greenacre M. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Softw.* 20: 1–13.

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

Nychka D, Furrer R, Sain S. 2013. fields: tools for spatial data. Available from: http://cran.r-project.org/web/packages/fields/index.html.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2013. vegan: community ecology package. R package version 2.0-3. Available from: http://cran.r-project.org/web/packages/vegan/index.html.

Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A. 2006. Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Sci Int.* 157: 23–35.

Ownby CP. 1988. Early Nguni history: the linguistic evidence and its correlation with archaeology and oral tradition [Ph.D. thesis]. [California (LA)]: University of California.

Patin E, Siddle KJ, Laval G, et al. 2014. The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat Commun.* 5:3163.

Pereira L, Gusmão L, Alves C, Amorim A, Prata MJ. 2002. Bantu and European Y-lineages in sub-Saharan Africa. *Ann Hum Genet.* 66: 369–378.

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet.* 65:439–458.

Phillipson DW. 2005. African archaeology. Cambridge (MA): Cambridge University Press.

Pickrell JK, Patterson N, Barbieri C, et al. 2012. The genetic prehistory of southern Africa. *Nat Commun.* 3:1143.

Pickrell JK, Patterson N, Loh PR, Lipson M, Berger B, Stoneking M, Pakendorf B, Reich D. 2014. Ancient west Eurasian ancestry in southern and eastern Africa. *Proc Natl Acad Sci U S A.* 111: 2632–2637.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967.

Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG, Behar DM. 2010. Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. *Am J Hum Genet.* 86:611–620.

Quintana-Murci L, Quach H, Harmant C, et al. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A.* 105:1596–1601.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105: 767–779.

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

Schlebusch CM, de Jongh M, Soodyall H. 2011. Different contributions of ancient mitochondrial and Y-chromosomal lineages in "Karretjie people" of the Great Karoo in South Africa. *J Hum Genet.* 56: 623–630.

Schlebusch CM, Lombard M, Soodyall H. 2013. MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. *BMC Evol Biol.* 13:56.

Sikora M, Laayouni H, Calafell F, Comas D, Bertranpetit J. 2011. A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations. *Eur J Hum Genet.* 19:84–88.

Sturrock K, Rocha J. 2000. A multidimensional scaling stress evaluation table. *Field Method.* 12:49–60.

Tesele TF. 1994. Symbols of power: beads and flywhisks in traditional healing in Lesotho [BA thesis]. [Cape Town (South Africa)]: University of Cape Town.

Thioulouse J, Chessel D, Dolédec S, Olivier JM. 1997. ADE-4: a multivariate analysis and graphical display software. *Stat Comput.* 7:75–83.

Tishkoff SA, Gonder MK, Henn BM, et al. 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol.* 24:2180–2195.

Tishkoff SA, Reed FA, Friedlaender FR, et al. 2009. The genetic structure and history of Africans and African Americans. *Science* 324: 1035–1044.

Turner FJ. 1962. The frontier in American history. New York: Henry Holt.

Underhill PA, Shen P, Lin AA, et al. 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet.* 26: 358–361.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 30:E386–E394.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.

Wadley L. 1996. Changes in the social relations of precolonial hunter–gatherers after agropastoralist contact: an example from the Magaliesberg, South Africa. *J Anthropol Archaeol.* 15:205–217.

Walker NJ. 1995. Late Pleistocene and Holocene hunter-gatherers of the Matopos: an archaeological study of change and continuity in Zimbabwe. Uppsala (Sweden): Societas Archaeologica Upsaliensis.

Wood ET, Stover DA, Ehret C, et al. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet.* 13:867–876.

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

Zalloua PA, Xue Y, Khalife J, et al. 2008. Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am J Hum Genet.* 82:873–882.