ARTICLE

Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people

Naser Ansari Pour^{*,1}, Christopher A Plaster¹ and Neil Bradman¹

The expansion of the Bantu-speaking people (EBSP) during the past 3000–5000 years is an event of great importance in the history of humanity. Anthropology, archaeology, linguistics and, in recent decades, genetics have been used to elucidate some of the events and processes involved. Although it is generally accepted that the EBSP has its origin in the so-called Bantu Homeland situated in the area of the border between Nigeria and the Grassfields of Cameroon, and that it followed both western and eastern routes, much less is known about the number and dates of those expansions, if more than one. Mitochondrial, Y-chromosome and autosomal DNA analyses have been carried out in attempts to understand the demographic events that have taken place. There is an increasing evidence that the expansion was a more complex process than originally thought and that neither a single demographic event nor an early split between western and eastern groups occurred. In this study, we analysed unique event polymorphism and short tandem repeat variation in non-recombining Y-chromosome haplogroups contained within the E1b1a haplogroup, which is exclusive to individuals of recent African ancestry, in a large, geographically widely distributed, set of sub-Saharan Africans (groups = 43, $n = 2757$), all of whom, except one Nilo-Saharan-speaking group, spoke a Niger-Congo language and most a Bantu tongue. Analysis of diversity and rough estimates of times to the most recent common ancestors of haplogroups provide evidence of multiple expansions along eastern and western routes and a late, exclusively eastern route, expansion.

European Journal of Human Genetics (2013) **21**, 423–429; doi:10.1038/ejhg.2012.176; published online 15 August 2012

Keywords: Africa; Bantu expansion; Y chromosome; E1b1a; TMRCA

INTRODUCTION

The early development of agriculture triggered significant population growth, resulting in the expansion of early farming populations, along with the spread of language families in many parts of the world, including Africa.¹ The many advantages of agricultural subsistence over foraging is a likely contributing factor to the rapid expansion of agriculturists and their languages during the holocene.² A well-known example of this phenomenon in Africa is the expansion of the Bantu-speaking people (EBSP), which is thought, on the basis of linguistic evidence, to have started around 5000 years ago³ in the region on the border between modern day eastern Nigeria and Cameroon.⁴ It is widely accepted that there was an early split into eastern and western routes in which farmers first expanded east and also, within 1500 years, reached West-Central Africa. After that the expansion is thought to have taken two directions with one wave moving along the south-western coast (West-Bantu route) and the other moving further east, forming the eastern Bantu core by 3000 years before present (YBP). Subsequently, the expansion is thought to have continued along the south-eastern coast (East-Bantu route).⁵ In an alternative model, the split came later after passage through the rain forest.^{3,4} The Bantu language family is distributed throughout most of sub-equatorial Africa and is the continent's largest, both in terms of the numbers of individuals speaking it and its geographic spread.^{2,6,7} This level of linguistic homogeneity among geographically distant

populations across sub-Saharan Africa supports the suggestion of rapid expansion. Because of the separation of the two major waves around 3500 YBP and the subsequent isolation of many groups, modern Bantu languages can be divided into West and East Bantu.^{8,9} Non-Bantu languages (that do not belong to the Niger-Congo phylum, eg, Khoisanid languages in the southwest of the continent and a few Cushitic and Nilo-Saharan tongues in the northeast⁶) have survived in areas that have not experienced extensive inward migration of Bantu speakers. It is also suggested that although the Bantu-speaking agriculturists may have replaced, to a substantial extent, hunter gatherers in their path, they have also, in some places, co-existed and interbred with the original inhabitants.²

Archaeological evidence suggests that the early expansion of proto-Bantu speakers was associated with pre-Iron Age farming technology and did not involve smelting metals.³ The first evidence of metallurgy south of the Sahara was found at Nok in Nigeria and is dated to no earlier than 2500 YBP.¹⁰ Therefore, it is possible that with the aid of the new technology, further expansions may have occurred after the first dispersal of farmers. Because the Bantu languages on the eastern route are more homogeneous than those on the western route,¹¹ it is reasonable to speculate that later expansions occurred mainly on the eastern route.

Early genetic studies of Bantu-speaking people were based on classical gene frequency data. Attempts were made to identify genetic

¹The Centre for Genetic Anthropology, Research Department of Genetics, Evolution and Environment, University College London, London, UK

*Correspondence: Dr N Ansari Pour, The Centre for Genetic Anthropology, Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. Tel: +44 20 73549163; Fax: +44 20 81816991; E-mail: naser.pour@ucl.ac.uk

Received 21 December 2011; revised 7 June 2012; accepted 6 July 2012; published online 15 August 2012

relationships among EBSP groups in the context of Africa as a whole^{10,11} (also see Supplementary Figure S1¹²). The major finding of these studies was that genetic distances (F_{ST}) among all EBSP groups are much less than the average F_{ST} among West-African and Nilo-Saharan groups, indicating a considerable level of homogeneity among EBSP groups. More recently, based on over 1300 autosomal markers, Tishkoff *et al*¹³ showed that Bantu-speaking groups exhibit a considerable level of genetic similarity, a finding which is in good agreement with earlier studies mentioned above. The EBSP impact on African demography has, over the past decade, also been studied by analysing paternal and maternal sex-specific genetic systems (non-recombining region of the Y chromosome (NRY) and mitochondrial DNA (mtDNA)). As both NRY and mtDNA genetic systems have smaller effective population sizes than autosomal markers, they are more prone to genetic drift^{14–16} and are therefore more likely to differ among groups than are autosomal markers. However, because each is, in effect, a single linked locus, interpreting observed differences among groups must be undertaken with a high level of caution. The control region of the mtDNA sequence, due to its high mutation rate, has been extensively used in examining the impact of EBSP on the genetic landscape of sub-Saharan Africa.^{5,17–19} It has been postulated that some mtDNA haplogroups (eg, L3b, L3e and L2a), based on their distribution in sub-Saharan Africa, are associated with the EBSP, whereas the presence of haplogroup L1c at high frequency in some populations on the western route is thought to be the result of assimilation of local female hunter gatherers.¹⁷ It has been suggested that because agriculturist men are more likely to marry local women rather than vice versa,^{15,16} the maternal genetic profile of Bantu-speaking groups is marked by considerable diversity. Despite this level of diversity, however, there is a high level of similarity between groups.²⁰

The increase in the rate of identification of slowly mutating NRY binary markers (ie, unique event polymorphisms (UEPs))^{21–23} has resulted in many studies designed to investigate the paternally mediated genetic relationships of sub-Saharan African populations. Scozzari *et al*²⁴ and Underhill *et al*²⁵ found UEP (M2 and its analogues such as DYS271G) present at high frequencies specifically in sub-Saharan Africa and suggested this marker as a signature of EBSP. Since then, this marker (now defining the E1b1a haplogroup) has been typed in many groups across sub-Saharan Africa^{19,26–28} and, without exception, all studies have shown that the majority of NRY types in Bantu-speaking groups belong to this haplogroup. Although sampling in most NRY studies of sub-Saharan Africa has, in the past, been quite limited in terms of geographic coverage and sample sizes, the distribution of this haplogroup is relatively well described in groups living along both the postulated western and eastern routes of the EBSP, as well as in Senegal²⁹ and Cameroon^{27,30} in West Africa. Interestingly, de Filippo *et al*³¹ recently reported differences in the frequencies of haplogroups E1b1a and E1b1a7 between Bantu and Non-Bantu Niger-Congo speakers. As the EBSP shows a clearer genetic legacy in the paternally inherited genetic system compared with mtDNA (evident from high and similar frequencies of E1b1a) in sub-Saharan Africa,³² it is possible that, as suggested by de Filippo *et al*,³¹ fine-scale E1b1a typing of Bantu-speaking communities throughout sub-Saharan Africa may add more structure to the geographic distribution of haplogroups. This, we hypothesise, may shed light on routes taken during their expansion.

Pakendorf *et al*⁷ in a recent review of the contribution made by molecular genetic analysis to the study of EBSP concluded that patrilocality and possibly polygyny may have contributed to NRY, but not mtDNA, association with linguistic affinity. They note that in

studies to date, Eastern African groups are greatly underrepresented but essential for investigating the direction of expansion. They further observe that the lack of genetic data makes it premature to reach sweeping conclusions concerning the EBSP.

In this study, we analyse, as did Alves *et al*,³³ both UEP and short tandem repeat (STR) (in this study restricted to NRY) to show that geographic frequency distributions and the time to the most recent common ancestors (TMRCA) of haplogroups, comprising haplogroup E1b1a in 43 sub-Saharan African groups ($n = 2757$) with diverse linguistic affiliations (Supplementary Figure S1), reveal multiple waves of expansion from West Africa, with a late expansion along the eastern route but not the western. As a consequence, this study makes an important contribution to filling the gap. Pakendorf *et al*⁷ identify and provide evidence of greater complexity in the process of the EBSP as suggested by Alves *et al*³³ and Montano *et al*.³⁴

MATERIALS AND METHODS

Samples

Buccal swabs were collected from males >18 years old unrelated at the paternal grandfather level but otherwise randomly selected from 43 groups across sub-Saharan Africa (Supplementary Table S1, samples from Ghana, Nigeria and Cameroon were included in Veeramah *et al* (2010)³⁵ and from South Africa in Thomas *et al* (2000)³⁶). These locations mainly cover West, Central-West, East, South-East and South Africa. All of the groups characterised in this study speak a Niger-Congo language, except for the Anuak in south-west Ethiopia who speak a Nilo-Saharan language. All buccal swabs were collected anonymously with appropriate ethical approval and informed consent. Sociological data were also collected from most individuals, including age, current residence, birthplace, self-declared cultural identity, first language, second language and (when available) religion of the individual, as well as similar information on the individual's father, mother, paternal grandfather and maternal grandmother. The samples were classified into groups primarily by cultural identity, first language spoken and then by place of collection. Where collections from a particular group were made in more than one location, locations are represented by averages of geographic coordinates. DNA from Congolese samples was extracted using the Gentra protein precipitation method (Gentra Systems, Minneapolis, MN, USA). Previously collected buccal-swab DNA samples from ethnic groups across sub-Saharan Africa were extracted by the standard phenol-chloroform method. The range and mean of sample sizes of the 43 groups are 25–118 and 63, respectively.

Y-chromosome typing

A combination of UEPs and STRs in the paternally inherited NRY was typed in eight Congolese groups ($n = 591$). The polymorphic markers are six STRs (DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393) and four UEPs (M191, U175, U290 and U181) characterising the E1b1a haplogroup, which is modal in most population groups within the area of the EBSP.²⁵ The four UEPs were typed using a tetra primer ARMS PCR method³⁷ with minor modifications. The outer and two inner fragments were amplified in a 10- μ l reaction volume containing 1 μ l (~ 1 ng) of template DNA, 1.6 μ l (50 μ M) dNTPs, 9.3 nM TaqStart monoclonal antibody (BD Biosciences Clontech, Oxford, UK), 0.13 U of Taq polymerase (HT Biotech, Cambridge, UK) and outer and inner primers (see Supplementary Table S2 for primer details). All samples (96-well plates) were then placed on a thermocycler under the following conditions: denaturation at 95 °C for 5 min, followed by 35 cycles of denaturation (95 °C) for 45 s, annealing (see Supplementary Table S2 for annealing temperatures) for 45 s and elongation (72 °C) for 45 s. The final step of the PCR programme was a 7-min extension at 72 °C before a 30 min hold at 4 °C. Where samples were ancestral for the four UEP markers, a further six to eleven UEPs (UEP1 and UEP2 kits: sY81, SRY4064, YAP, SRY10831, M13, M9, SRY465, M20, Tat, 92R7 and M17) were typed.³⁸ NRY haplogroups were classified according to the nomenclature of the Y-Chromosome Consortium³⁹ (Figure 1) and STR repeat sizes were assigned according to the nomenclature of Kayser *et al*.⁴⁰ Additionally, the four E1b1a-specific UEPs were typed in

Figure 1 Genealogical relationships of UEP markers used to define NRY haplogroups. The box identifies the E1b1a clade, exclusively observed in population groups with recent African ancestry.

1820 samples, previously characterised as E1b1a in the TCGA database (published^{35,36} and unpublished data), from the 35 non-Congo, sub-Saharan groups listed in Supplementary Table S1. Although the battery of the NRY markers typed in UEP kits gives a relatively crude resolution of NRY haplogroups, the typing of four UEP markers within E1b1a considerably increases the resolution of NRY types associated with EBSP.³²

Statistical analysis

Haplotype diversity, h , and its SE were estimated from unbiased formulae of Nei⁴¹ and was performed using Arlequin software version 3.0.⁴² Average squared difference (ASD) in STR allele size between all chromosomes and the presumed ancestral haplotype (assumed to be the modal haplotype), averaged over loci, were estimated using YTIME software,⁴³ and corresponding 95% confidence intervals were calculated as described in Thomas *et al*⁴⁴ using the 'R' environment of statistical computing (www.R-project.org). The TMRCA was estimated using an average NRY STR mutation rate of 0.002⁴⁵ and generation time of 25 years. The Fisher's exact test was also performed in the R environment. The probability of observing a particular haplotype, if present, in a randomly collected set was assessed by the equation $(1-q)^n = (1-P)$, where P is the probability of observing the haplotype, q is the minimum frequency of the haplotype to be observed and n is the number of chromosomes. According to the equation, the minimum frequency at which a haplotype is present for it to have a 95% probability of being observed, given that n chromosomes are typed, is $q = 1 - 10^{(\log(0.05)/n)}$.

RESULTS

Of the possible 17 haplogroups, 12 were observed in the complete data set with haplogroup E1b1a modal (0.847, range in population groups 0.389–0.957), both overall and in every sub-Saharan African group. Only two other haplogroups exceeded 5% of the total: BT* (xDE,KT) (7.5%) and E* (xE1b1a) (5.1%). Table 1 reports the frequencies of all observed haplogroups, including the component haplogroups of E1b1a. Haplogroup E1b1a7 or E1b1a8* is modal in all groups with the exception of Bankim (Cameroon) and Fante

(Ghana). The pooled frequencies of E1b1a component haplogroups, based on their geographic locations, are also shown in Figure 2.

All haplogroups within E1b1a were observed in the Bantu Homeland, West-Central Africa, East Africa and Ghana, whereas haplogroup E1b1a8a1a, although present in the Bantu Homeland and East Africa, was not observed in either Ghana or West-Central Africa.

Diversity (h) of E1b1a was calculated at the five component-haplogroup level ranged from 0.379 to 0.753, excluding the Anuak ($h = 0$). For comparison, the NRY haplotype diversity treating E1b1a as a single haplogroup ranged from 0.821 to 0.945, with the exception of Anuak who displayed a much lower diversity ($h = 0.516$).

The EBSP six-STR haplotype was modal in 36 out of the 43 groups (see Supplementary Table S3) and was almost always a member of E1b1a8 (frequency of 96.4%, $P < 0.0001$). Table 2 contains the six-STR haplotype gene diversities for E1b1a component haplogroups present in all three West, West-Central and East-Central regions.

The TMRCA for each haplogroup-defining UEP (with at least 20 chromosomes) is presented in Table 3 along with regions and countries within which each haplogroup was observed. The finer branches of the genealogical tree were associated with lower estimates of TMRCA (Figure 1). TMRCA for E1b1a as a whole was estimated at 6175–6588 YBP with the TMRCA for the youngest haplogroup (E1b1a8a1a) estimated at 1100–1638 YBP.

DISCUSSION

We analyse frequencies of haplogroups and estimates of TMRCA to answer two questions: (a) Is there evidence of more than one 'expansion' of paternal line ancestors of Bantu-speaking people living in present day sub-Saharan Africa? and (b) If so, did those 'expansions' take different routes? We define 'expansion' in this context to mean diffusion of alleles. In doing so, we assume (a) that the NRY has a genealogy that, at least in that part of the genealogical tree analysed in this paper, can be unambiguously constructed using UEP polymorphisms⁴⁷ (Figure 2) and (b) ASD is a measure of STR diversity that increases linearly over time and that calculating ASD from the common ancestor of a random sample of NRY that are members of a haplogroup provides an estimate of the TMRCA.⁴³ Consistent with previous studies, we observed a high frequency modal of six-STR NRY haplotype (DYS19, 388, 390, 391, 392, 393:15–12–21–10–11–13) throughout the area of the EBSP.^{26,35,36} Interpreting the frequencies of the component haplogroups of E1b1a within the context of their geographic distribution and TMRCA values throws additional light on the expansions associated with the EBSP. These data are consistent with multiple expansion events southwards from West Africa. Haplogroup E1b1a7 (defined by M191) is modal in most groups in countries from Ghana to Mozambique and only at slightly lower frequency in South African Bantu speakers (33.8% compared with E1b1a8* at 37.8%). The TMRCA at 4700–5300 YBP is entirely consistent with the haplogroup being present in West Africa at the dawn of the EBSP. It is likely to have expanded south as the demographic events comprising the EBSP took place. The haplogroup E1b1a8, defined by U175, has a TMRCA of only 1863–2163 YBP but a geographic distribution, excepting the Anuak of Ethiopia, which is equally extensive as that of E1b1a7. If it is assumed that an earlier expansion had already taken place, this would be consistent with a subsequent, rapid expansion from West Africa southwards along both the western and eastern routes. This is consistent with the analysis of de Filippo *et al*,³¹ which is also supportive of a rapid expansion. It is interesting to speculate on the possibility that this later expansion was associated with the contemporaneous development of metallurgy.

Table 1 Frequency distribution of NRY haplogroups in 43 sub-Saharan African population groups

NRY UEP haplogroup
(according to the
nomenclature proposed
by Karafet *et al*¹⁸)

	<i>n</i>	<i>E1b1a*</i>	<i>E1b1a7</i>	<i>E1b1a8*</i>	<i>E1b1a8a1*</i>	<i>E1b1a8a1a</i>	<i>E*(x E1b1a)</i>	<i>DE*(x E)</i>	<i>BT*(x DE,KT)</i>	<i>K*(x L,N1c,O2b,P)</i>	<i>A3b2</i>	<i>P*(x R1a)</i>	<i>Y*(x BT,A3b2)</i>
Bembe (BE)	109	0.037	0.321	0.211	0.284		0.110				0.018	0.018	
Kuni (KU)	68	0.015	0.515	0.162	0.176		0.074		0.044				0.015
Lari (LA)	62		0.484	0.177	0.194		0.113		0.032				
Mboshi (MB)	91		0.440	0.286	0.187				0.044			0.011	0.033
Sundi (SU)	25	0.040	0.600	0.200	0.040		0.080		0.040				
Teke (TE)	63		0.429	0.222	0.159		0.079		0.063			0.032	0.016
Vili (VI)	108	0.056	0.500	0.250	0.139		0.019		0.028			0.009	
Yombe (YO)	65	0.031	0.277	0.369	0.138		0.092		0.046			0.015	0.031
Congo total	591	0.024	0.430	0.239	0.181		0.066		0.034		0.003	0.012	0.012
Bankim (BA)	33	0.364	0.182	0.152	0.152		0.061		0.061			0.030	
Foumban (FO)	117	0.068	0.667	0.068	0.051		0.043		0.085		0.009		0.009
Wum (WU)	116	0.026	0.621		0.310				0.043				
Cameroon total	266	0.086	0.586	0.049	0.177		0.026		0.064		0.004	0.004	0.004
Asante (AS)	94	0.245	0.309	0.106	0.245		0.043						0.053
Ewe (EW)	88	0.114	0.420	0.102	0.273		0.045		0.023			0.023	
Fante (FA)	60	0.300	0.250	0.167	0.217		0.017					0.033	0.017
Ghana total	242	0.211	0.335	0.120	0.248		0.037		0.008			0.017	0.025
Chewa (CH)	92	0.043	0.239	0.293	0.130	0.054	0.130		0.076	0.011	0.022		
Tumbuka (TU)	61	0.033	0.410	0.230	0.082	0.033	0.098		0.082	0.016	0.016		
Yao (YA)	56	0.071	0.321	0.250	0.054	0.036	0.071		0.196				
Malawi total	209	0.048	0.311	0.263	0.096	0.043	0.105		0.110	0.010	0.014		
Abak (AB)	56	0.054	0.357	0.107	0.161	0.232	0.036		0.054				
Afaha Eket (AE)	48	0.104	0.354	0.042	0.125	0.208	0.042		0.125				
Afaha Okpo (AO)	50	0.020	0.460	0.060	0.080	0.280	0.060	0.020	0.020				
Afaha Ukwong (AU)	49	0.041	0.408	0.102	0.143	0.163	0.020		0.102				0.020
Awa-Onna (AW)	28		0.500		0.179	0.107	0.036		0.179				
Calabar (CA)	99	0.061	0.475	0.121	0.101	0.141	0.010	0.010	0.071				0.010
Ediene Ikono (ED)	48	0.083	0.479	0.042	0.083	0.125	0.063		0.125				
Efut Akpabuyo (EF)	48	0.125	0.438	0.042	0.042	0.125	0.021		0.188			0.021	
Efut Odukpani (EO)	49	0.020	0.449	0.143	0.082	0.204	0.041		0.061				
Ejagham-Akamkpa (EA)	47	0.021	0.426	0.085	0.277	0.043	0.043		0.085			0.021	
Ejagham-Calabar (EC)	83	0.024	0.518	0.048	0.193	0.108	0.024		0.060			0.024	
Eziagu Nenwe (EZ)	49	0.286	0.469	0.020	0.102	0.082			0.041				
Ikono (IK)	42		0.429	0.167	0.143	0.071	0.024		0.167				
Itam (IT)	50	0.080	0.520	0.060	0.120	0.160			0.040				0.020
Nike Enugu (NE)	54	0.019	0.685		0.074	0.130	0.019	0.019	0.056				
Nnung Ndem-Onna (NN)	48	0.042	0.563	0.063	0.250	0.021	0.021	0.021	0.021				
Nsit (NS)	36	0.028	0.361	0.111	0.250	0.167	0.028		0.056				
Ntan Ibiono (NT)	50	0.040	0.380	0.160	0.140	0.180	0.020		0.080				
Obong Itam (OB)	50		0.480	0.060	0.140	0.120	0.040	0.020	0.100	0.020		0.020	
Oku-Itu (OI)	49		0.510	0.102	0.102	0.163			0.122				
Oku-Uyo (OU)	48	0.042	0.396	0.083	0.208	0.167	0.063		0.042				
Ukpom Ete (UE)	50	0.060	0.480	0.140	0.100	0.080	0.040	0.020	0.080				
Uwanse (UW)	50	0.060	0.460	0.060	0.060	0.240			0.100				0.020
Nigeria total	1181	0.053	0.464	0.080	0.135	0.145	0.027	0.005	0.082	0.001		0.004	0.003
Bantu speakers- Pretoria (BN)	98	0.061	0.255	0.286	0.071	0.082	0.051		0.153	0.010			0.031
Sena-Mozambique (SE)	62	0.161	0.306	0.274	0.032	0.016	0.129		0.081				
Anuak-Ethiopia (AN)	108		0.389				0.167		0.250		0.194		

Modal haplogroups are shown in bold type. Haplogroups unobserved in any sample are not shown. *n* is sample size.

Figure 2 Visual representation of the distribution of E1b1a component haplogroups in sub-Saharan African groups with sample totals. Sectors in pie charts are coloured according to the haplogroup colour code to the left. Sample sizes are indicated within the pie charts. Samples in the Congolese data set have been divided into three pie charts representing Bantu H, B and C speakers.

Table 2 STR haplotype diversity within E1b1a component haplogroups present in all Bantu-speaking groups

Haplogroup	Marker	Haplotype diversity, <i>h</i> (SD)		
		West Africa	West-Central Africa	East-Central Africa
				Malawi and
				Mozambique
			Congo (Guthrie H,	
			B and C Bantu	(Guthrie Bantu N
			languages)	and P languages)
		Nigeria and		
		Cameroon		
E1b1a7	M191	0.934(0.003)	0.875(0.010)	0.907(0.016)
E1b1a8	U175	0.880(0.019)	0.799(0.026)	0.789(0.040)
E1b1a8a1	U290	0.713(0.032)	0.584(0.052)	0.632(0.104)

Abbreviations: *h*, Nei's unbiased estimator of gene diversity; *n*, sample size; NA, not applicable; STR, short tandem repeat.

Because E1b1a* is a paragroup and making inferences based on its distribution may be misleading,⁴⁶ STR gene-diversity values were not calculated.

The distribution of haplogroup E1b1a8a1* defined by U290 in the absence of U181 with a TMRCA of 1413–1725 YBP is similar to that of E1b1a8 and may be interpreted in the same way. The distribution of haplogroup E1b1a8a1a (defined by U181) with a very recent TMRCA of only 1100–1638 YBP is very different, however, being restricted to Nigeria and the east side of sub-Saharan Africa (Figure 2). As a consequence it is consistent with a late, rapid expansion from south of the Grassfields of Cameroon that did not include expansion along the earlier western route. Because the West-Central African E1b1a data set is sufficiently large ($n=516$; eight

groups), we would have expected to observe the E1b1a8a1a haplotype, if present at a frequency as low as 0.0058. Therefore, it is unlikely that the absence of this haplogroup is due to drift after the initial stage of expansion when only a small number of individuals may have been involved or is simply not being observed in the present study. We note that the phenomenon of surfing can explain the absence of an allele in only some groups that are the consequence of range expansion.^{48,49} However, unless the allele (in this case NRY belonging to haplogroup E1b1a8a1a) became extinct early in the western route expansion (which is, in effect, the same as not having been part of that expansion), there is no reason to suppose that extinction of the haplogroup in western route groups (Guthrie classification H, B and C) was more likely than in eastern groups (Guthrie classification N and P). See Supplementary Table S4 for Guthrie classifications of all Bantu-speaking groups included in the analysis. In this study, haplogroup E1b1a8a1a, the haplogroup with the shortest TMRCA, was observed in all eastern data sets (three from Malawi, one from Mozambique (in both cases, all speakers of Guthrie classification Bantu languages N and P spoken on the eastern side of Africa) and one from Pretoria, n (samples) = 18) but in none of the eight western groups (all speakers of Guthrie classification Bantu languages H, B and C spoken on the western side of Africa) (Fisher's exact test: haplogroup present/absent in data set $P=0.0008$; haplogroup frequency $P<0.0001$). Comparisons made without including data sets from South Africa and Mozambique, so as to exclude the possibility of admixture between western and eastern Bantu-speaking expansions in the southern extremity of the continent, remain significant for both presence/absence of E1b1a8a1a in data sets and for frequency of the haplogroup ($P<0.01$). The genetic data are thus in broad agreement with analysis based on linguistic studies, which suggests that the spread of Bantu languages is the consequence of successive dispersals and that a single large-scale migration by Bantu speakers is unlikely.³ It is also consistent with suggestions that differences between eastern and western Bantu languages are a consequence of expansion patterns.³ This interpretation suggests the absence of substantial male-mediated gene flow from East-Central Africa to West-Central Africa during the past millennium, because had it occurred, it would be expected that examples of haplogroup E1b1a8a1a would have been observed in the Congolese groups included in this study. Further support for the EBSP origin from the Nigeria/Cameroon area comes from the observation that E1b1a component-haplogroup STR diversities are greater in West Africa than in either West-Central or East-Central Africa (Table 2).

Recently, Alves *et al*³³ analysing a battery of 14 DIPSTRs (ie, deletion/insertion polymorphisms tightly linked to STRs) in 19 Bantu-speaking groups from Mozambique and Angola concluded that it is becoming increasingly difficult to accept models, suggesting an early split between eastern and western Bantu-speaking populations, whereas Montano *et al*³⁴ analysing NRY UEPs and STRs in groups from Nigeria, Cameroon, Gabon and Congo concluded that the evolutionary scenario is more complex than previously thought. Our analysis of NRY from groups over a wide geographic area is consistent with both these conclusions.

We conclude that analysis of NRY in 43 widely distributed population groups from across sub-Saharan Africa provides evidence of multiple expansions from West Africa along the western and eastern routes and a late specifically eastern expansion at some time during the past two millennia during a period in which male-mediated gene flow from East-Central to West-Central Africa does not appear to have taken place, at least to any significant extent. Future studies that examine variation in the NRY E1b1a clade in

Table 3 Estimated TMRCA of E1b1a UEP dates and distribution of E1b1a component haplogroups in sub-Saharan Africa

Haplogroup	Marker	n	ASD (95% CI)	TMRCA (YBP)	Geographic presence (Y/N)					Total
					West Africa		West-Central Africa	East Africa	South-East Africa	
					Ghana	Nigeria	Cameroon	Congo	Ethiopia	
BT*(xDE,KT)	SRY10831	206	1.036(0.939–1.138)	11 731–14 229	Y	Y	Y	Y	Y	Y
A3b2	M13	27	0.556(0.401–0.735)	5015–9182	N	N	Y	Y	Y	Y
E*(E1b1a)	SRY4064	140	0.920(0.816–1.036)	10 193–12 946	Y	Y	Y	Y	Y	Y
E1b1a	sY81	2336	0.510(0.494–0.527)	6175–6588	Y	Y	Y	Y	Y	Y
E1b1a7	M191	1190	0.404(0.384–0.424)	4800–5300	Y	Y	Y	Y	Y	Y
E1b1a8	U175	969	0.161(0.149–0.173)	1863–2163	Y	Y	Y	Y	N	Y
E1b1a8a1	U290	591	0.125(0.113–0.138)	1413–1725	Y	Y	Y	Y	N	Y
E1b1a8a1a	U181	189	0.109(0.088–0.131)	1100–1638	N	Y	N	N	N	Y

Abbreviations: CI, confidence interval; n, sample size; N, no; Y, yes.

Bantu-speaking population groups representing the East African coast will help to further elucidate the late eastern EBSP.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all DNA donors and those assisting in sample collection and Professor Mark Thomas and Dr Krishna Veeramah for their support with typing and helpful comments and suggestions on the manuscript. NAP was supported by NERC-Case PhD studentship.

- Bellwood P: Early agriculturalist population diasporas? Farming, languages, and genes. *Annu Rev Anthropol* 2001; **30**: 181–207.
- Diamond J, Bellwood P: Farmers and their languages: the first expansions. *Science* 2003; **300**: 597–603.
- Vansina J: New Linguistic Evidence and 'the Bantu Expansion'. *J Afr Hist* 1995; **36**: 173–195.
- Newman JL: *The Peopling of Africa: A Geographic Interpretation*. New Haven: Yale University Press, 1995.
- Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A: Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* 2001; **65**: 439–458.
- Nurse D: Bantu languages; in Brown K (ed) *Encyclopedia of Language and Linguistics*. Oxford: Elsevier Ltd, 2006; pp 679–685.
- Pakendorf B, Bostoen K, de Filippo C: Molecular perspectives on the Bantu expansion: a synthesis. *Lang Dyn Change* 2011; **1**: 50–88.
- Holden CJ: Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proc R Soc Lond B* 2002; **793**: 793–799.
- Jobling MA, Hurler ME, Tyler-Smith C: Human Evolutionary Genetics: Origins. *People and Disease*. Abingdon: Garland Science, 2004.
- Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. New Jersey: Princeton University Press, 1994.
- Excoffier L, Pellegrini A, Langaney A: Genetics and history of sub-Saharan Africa. *Am J Phys Anthropol* 1987; **30**: 151–194.
- Lewis MP: *Ethnologue: Languages of the World*. Dallas: SIL International, 2009.
- Tishkoff SA, Reed FA, Friedlaender FR *et al*: The genetic structure and history of Africans and African Americans. *Science* 2009; **324**: 1035–1044.
- Roewer L, Kayser M, de Knijff P *et al*: A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Sci Int* 2000; **114**: 31–43.
- Destro-Bisol G, Donati F, Coia V *et al*: Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol Biol Evol* 2004; **21**: 1673–1682.
- Wood ET, Stover DA, Ehret C *et al*: Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet* 2005; **13**: 867–876.
- Salas A, Richards M, De la FT *et al*: The making of the African mtDNA landscape. *Am J Hum Genet* 2002; **71**: 1082–1111.
- Salas A, Richards M, Lareu MV *et al*: The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* 2004; **74**: 454–465.
- Beleza S, Gusmao L, Amorim A *et al*: The genetic legacy of western Bantu migrations. *Hum Genet* 2005; **117**: 366–375.
- Castri L, Tofanelli S, Garagnani P *et al*: 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* 2009; **140**: 302–311.
- Hammer MF: A recent common ancestry for human Y chromosomes. *Nature* 1995; **378**: 376–378.
 - Underhill PA, Jin L, Lin AA *et al*: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; **7**: 996–1005.
 - Underhill PA, Shen P, Lin AA *et al*: Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000; **26**: 358–361.
 - Scozzari R, Cruciani F, Santolamazza P *et al*: Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 1999; **65**: 829–846.
 - Underhill PA, Passarino G, Lin AA *et al*: The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001; **65**: 43–62.
 - Pereira L, Gusmao L, Alves C *et al*: Bantu and European Y-lineages in sub-Saharan Africa. *Ann Hum Genet* 2002; **66**: 369–378.
 - Cruciani F, Santolamazza P, Shen P *et al*: A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 2002; **70**: 1197–1214.
 - Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J: On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol Biol* 2009; **9**: 80.
 - Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA: Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 2002; **70**: 265–268.
 - Luis JR, Rowold DJ, Regueiro M *et al*: The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 2004; **74**: 532–544.
 - de Filippo C, Barbieri C, Whitten M *et al*: Y-chromosomal variation in sub-Saharan Africa: Insights into the history of Niger-Congo groups. *Mol Biol Evol* 2011; **28**: 1255–1269.
 - Berniell-Lee G, Calafell F, Bosch E *et al*: Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol* 2009; **26**: 1581–1589.
 - Alves I, Coelho M, Gignoux C, Damasceno A, Prista A, Rocha J: Genetic homogeneity across Bantu-speaking groups from Mozambique and Angola challenges early split scenarios between East and West Bantu populations. *Hum Biol* 2011; **83**: 13–38.
 - Montano V, Ferri G, Marcari V *et al*: The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa. *Mol Ecol* 2011; **20**: 2693–2708.
 - Veeramah KR, Connell BA, Ansari Pour N *et al*: 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 Evol Biol* 2010; **10**: 92.
 - Thomas MG, Parfitt T, Weiss DA *et al*: Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba – the 'Black Jews of Southern Africa'. *Am J Hum Genet* 2000; **66**: 674–686.
 - Ye S, Dhillon S, Ke X, Collins AR, Day IN: An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 2001; **29**: E88.
 - Thomas MG, Bradman N, Flinn HM: High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. *Hum Genet* 1999; **105**: 577–581.
 - Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 2008; **18**: 830–838.
 - Kayser M, Caglia A, Corach D *et al*: Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 1997; **110**: 125–129.
 - Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press, 1987.
 - Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 2005; **1**: 47–50.
 - Behar DM, Thomas MG, Skorecki K *et al*: Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. *Am J Hum Genet* 2003; **73**: 768–779.

- 44 Thomas MG, Skorecki K, Ben Ami H, Parfitt T, Bradman N, Goldstein DB: Origins of old testament priests. *Nature* 1998; **394**: 138–140.
- 45 Gusmao L, Sanchez-Diz P, Calafell F *et al*: Mutation rates at Y chromosome specific microsatellites. *Hum Mutat* 2005; **26**: 520–528.
- 46 Weale ME, Shah T, Jones AL *et al*: Rare deep-rooting Y chromosome lineages in humans: lessons for phylogeography. *Genetics* 2003; **165**: 229–234.
- 47 Trombetta B, Cruciani F, Sellitto D, Scozzari R: A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. *PLoS ONE* 2011; **6**: e16073.
- 48 Edmonds CA, Lillie AS, Cavalli-Sforza LL: Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 2004; **101**: 975–979.
- 49 Klopstein S, Currat M, Excoffier L: The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* 2006; **23**: 482–490.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)