Mosaic maternal ancestry in the Great Lakes region of East Africa

Verónica Gomes,[1,2] Maria Pala,[*,3] António Salas,[*,4] Vanesa Álvarez-Iglesias,[4] António Amorim,[1,2,5] Alberto Gómez-Carballa,[4] Ángel Carracedo,[4] Douglas J. Clarke,[3] Catherine Hill,[3] Maru Mormina,[6,7] Marie-Anne Shaw, [6,8] David W. Dunne,[9] Rui Pereira,[1,2] Vânia Pereira ,[10] Maria João Prata,[1,2,5] Paula Sánchez-Diz,[4] Teresa Rito,[11,12] Pedro Soares,[*13] Leonor Gusmão,[+,1,2,14] Martin B. Richards.[+,3,6]

[1]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal.

[2]Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal.

[3]Department of Biological Sciences, School of Applied Sciences, University of Huddersfield, UK

[4]Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses and Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, CIBERER, Galicia, Spain

[5]Faculty of Sciences, University of Porto, Portugal

[6]Faculty of Biological Sciences, University of Leeds, UK

[7]Department of Applied Social Studies, University of Winchester, UK

[8]Leeds Institute of Molecular Medicine, Faculty of Medicine and Health, University of Leeds, UK

[9]Department of Pathology, University of Cambridge, UK

[10]Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[11]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal

[12]ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal

[13]Centre of Molecular and Environmental Biology, University of Minho, Braga, Portugal

[14]DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil.

[*, +] These authors contributed equally to this work.

Corresponding Author:
Martin B. Richards
Department of Biological Sciences
School of Applied Sciences,
University of Huddersfield,
Queensgate,
Huddersfield, HD1 3DH, UK
Tel: +44-1484 471676;
Email: m.b.richards@hud.ac.uk

Abstract

The Great Lakes lie within a region of East Africa with very high human genetic diversity, home of many ethno-linguistic groups usually assumed to be the product of a small number of major dispersals. However, our knowledge of these dispersals relies primarily on the inferences of historical linguistics and oral traditions, with attempts to match up the archaeological evidence where possible. This is an obvious area to which archaeogenetics can contribute, yet Uganda, at the heart of these developments, has not been studied for mitochondrial DNA (mtDNA) variation. Here, we compare mtDNA lineages at this putative genetic crossroads across 409 representatives of the major language groups: Bantu speakers and Eastern and Western Nilotic speakers. We show that Uganda harbours one of the highest mtDNA diversities within and between linguistic groups, with the various groups significantly differentiated from each other. Despite an inferred linguistic origin in South Sudan, the data from the two Nilotic-speaking groups point to a much more complex history, involving not only possible dispersals from Sudan and the Horn but also large-scale assimilation of autochthonous lineages within East Africa and even Uganda itself. The Eastern Nilotic group also carries signals characteristic of West-Central Africa, primarily due to Bantu influence, whereas a much stronger signal in the Western Nilotic group suggests direct West-Central African ancestry. Bantu speakers share lineages with both Nilotic groups, and also harbour East African lineages not found in Western Nilotic speakers, likely due to assimilating indigenous populations since arriving in the region ~3000 years ago.

Keywords: Nilotic languages, Bantu dispersals, East Africa, mitochondrial DNA, phylogeography

Introduction

Eastern Africa has a central place in the evolution of anatomically modern humans for at least two reasons. Firstly, it is considered by many to be the region from which modern humans spread from Africa into the rest of the world, approximately 60,000 years ago (60 ka) (e.g. Atkinson et al. 2008; Gonder et al. 2007; Macaulay et al. 2005; Tishkoff et al. 1996; Watson et al. 1997), although some paleoanthropologists and archaeologists favour North Africa [see discussion in Balter (2011)]. Secondly, it has been argued by some to be the place of origin of modern *Homo sapiens*. Whilst the main line of evidence is that the earliest transitional/modern fossil skeletal remains have been found in Ethiopia (McDougall et al. 2005; White et al. 2003), Eastern Africa also harbours deep branches in the mtDNA and Y-chromosome phylogenies, and genetic diversity estimated here is often higher than anywhere else in the world (Gonder et al. 2007; Hassan et al. 2008; Pagani et al. 2012; Poloni et al. 2009; Salas et al. 2002) – although Central Africa has become an increasingly plausible candidate (Cruciani et al. 2011; Rito et al. 2013). On the other hand, Tishkoff et al. (2009) suggested, on the basis of genome-wide autosomal studies, that Southwest Africa, rather than Eastern Africa, might be the most likely place for the geographical origin of modern humans within the continent. A similar conclusion was reached by Henn et al. (2011), who compared F_{ST} distances and linkage disequilibrium (LD) patterns in hunter-gatherer populations from both Eastern and Southern Africa with other African populations, finding the lowest LD values and highest F_{ST} values amongst the Southern African groups.

An alternative explanation for the high genetic diversity in Eastern Africa might (at least in part) be admixture between a number of genetically divergent human populations (Hassan et al. 2008; Poloni et al. 2009). For example, not only has Ethiopia experienced both ancient and recent gene flow from Eurasia (Kivisild et al. 2004; Olivieri et al. 2006; Pagani et al. 2012; Richards et al. 2003), but Eastern Africa in general exhibits high environmental, economic, social, cultural and linguistic diversity, as illustrated, for instance, by a number of surviving foraging populations in Tanzania (Campbell and Tishkoff 2010). The Great Lakes region in particular has, on the evidence of historical linguistics and archaeology, been the meeting point of diverse African groups for millennia (Maxon 2009; Newman 1995; Phillipson 2005; Poloni et al. 2009; Tishkoff et al. 2009).

In East Africa, present-day Uganda was part of the territory located on the fringe of the 'Bantu expansion' that brought presumed Bantu-speaking groups carrying the Chifumbaze material culture complex (including iron-working) from ~3 ka, occupying much of the most productive lakeside and riverine agricultural land (Phillipson 2005). However, Uganda has also been crossed and occupied by dispersals of Nilotic speakers (Maxon 1994; Newman 1995; Pazzaglia 1982). It is assumed that, from at least ~3 ka, Bantu- and Nilotic-speaking groups began assimilating the indigenous populations previously inhabiting East Africa, most likely including speakers of Afroasiatic Southern Cushitic, Nilo-Saharan Central Sudanic and possibly also Khoisan languages.

The Nilotic languages belong to the Southern branch of the Eastern Sudanic family of Nilo-Saharan languages, although there is considerable disagreement about the branching within Nilo-Saharan (Bender 2000; Ehret 1998, 2001). The ancestry of Nilotic-speaking populations in the Great Lakes region has been reconstructed by attempting to triangulate and combine evidence from oral history, historical linguistics and archaeology (e.g. Newman 1995), but tying in the latter is difficult (Ambrose 1982; Ehret 1998; Phillipson 2005), ethnicities have usually been highly porous (Kusimba and Kusimba 2005) and the role of language shift may have been considerable (Ehret 1998).

They lie at the southernmost extreme of the geographic range of the Nilo-Saharan phylum and it has been proposed (on linguistic grounds) that they had arisen by the beginning of the first millennium BC, somewhere in the vicinity of the Nile swamps of the Bahr-el Ghazal in Southern Sudan, subsequently expanding into East African territory already populated by stone-age cereal agriculturalists speaking a variety of Sudanic and Cushitic languages, and later on also iron-using Bantu speakers (David 1982; Ehret 1998, 2001).

Three branches are recognized: Southern, Western, and Eastern Nilotic. It is widely thought that the Southern group spread south into East Africa earlier, possibly more than 3 ka, before the arrival of Bantu speakers. Further wave of dispersals in the late first millennium AD, of either Southern or Eastern Nilotic speakers, may have accompanied the spread of the Later Iron Age "rouletted" ceramics that replaced the Urewe ware associated with the Bantu Early Iron Age (Ambrose 1982; David 1982; Oliver 1982). This would have spread the agro-pastoral way of life throughout Bantu Africa, leading to the occupation of more arid territories on a much larger scale, although this is less widely accepted (Phillipson

2005). The Teso-Turkana branch of Eastern Nilotic in Kenya is thought to have split off by 100–500 AD (Ambrose 1982; Newman 1995). However, Eastern Nilotic speakers are thought to have settled substantially in Eastern Uganda only since the seventeenth century AD (Maxon 2009; Newman 1995). The Western Nilotic expansion (of Lwo speakers) is also thought to have taken place over the last five centuries (Newman 1995). These dispersals established an intricate and complex mosaic of relationships between the different ethnic groups.

Most previous genetic studies of sub-Saharan Africa have mainly focused either on speakers of click-languages and other hunter-gatherer populations (Batini et al. 2007; Behar et al. 2008; Gonder et al. 2007), or on the Bantu dispersals (e.g. Batai et al. 2013; Beleza et al. 2005; Castri et al. 2009; Coia et al. 2005; Gonder et al. 2007; Plaza et al. 2004; Quintana-Murci et al. 2008; Richards et al. 2004; Salas et al. 2002; Scozzari et al. 1994). Several also studied populations speaking Nilo-Saharan languages (Krings et al. 1999; Poloni et al. 2009; Tishkoff et al. 2007; Tishkoff et al. 2009; Watson et al. 1996; Wood et al. 2005) and Afroasiatic languages (Boattini et al. 2013), and a substantial genome-wide dataset from Ethiopia points to significant differentiation amongst speakers of the different language phyla (Pagani et al. 2012). Studies of mtDNA in Bantu and Nilotic-speaking groups in Kenya have shown that they have much in common, but that they can nevertheless be differentiated phylogeographically by language phylum (Castri et al. 2008), and that even distinct Bantu speaking groups can differ significantly genetically, possibly depending on the extent of assimilation of local lineages (Batai et al. 2013). Genome-wide studies suggest extensive contact and gene flow between Nilo-Saharan and Cushitic speakers, with an ancestral cluster attributed to Nilo-Saharan speaking groups most common in Central/Southern Sudan and decreasing in frequency in Nilo-Saharan speakers moving south, consistent with a model of dispersal with increasing assimilation. There is a substantial level of a possible “Cushitic” substrate in Nilo-Saharan-speaking populations of Kenya and Tanzania, also present to a lesser extent in Bantu speakers. Although there is evidence for a significant West-Central African component in Bantu speakers (Batai et al. 2013), there is also evidence for exchange/assimilation between Bantu and Nilotic speakers in both regions (Tishkoff et al. 2007).

The present study is the first to focus specifically on Ugandan populations comparing mtDNA lineages across the major linguistic communities. Each has been suggested, to have originated outside this region (Ehret 1998, 2001; Phillipson 2005), but the relative extent of dispersal *versus* acculturation in each, and the level of interaction between them, has yet to be assessed. Here, within the context of more 8000 comparative sequences from across Africa, we begin such an assessment by characterizing their maternal lineages, as a crucial first step in reconstructing in detail the history of the populations of the region.

Materials and Methods (Summary)

We refer to Kenya, Tanzania and Uganda as East Africa, and to the broader region including East Africa, the Horn of Africa and South Sudan as Eastern Africa. A total sample of 409 healthy individuals from Uganda participated in this work, under informed consent. 109 were from Bantu and 300 from Nilo-Saharan speakers, the latter including 290 Nilotic (243 Eastern and 47 Western Nilotic) and 10 Central Sudanic speakers (Table S1).

For 261 Ugandan samples we analysed the entire mtDNA control region (16024–16569; and 1–576) and for the remaining 148 samples the HVS-I region (16011–16497; minimum extent 16085–16430). The classification of the mtDNA sequences into haplogroups followed PhyloTree, mtDNA Build 16, 19 Feb 2014 (van Oven and Kayser 2009).

We compared the population samples from Uganda with published data pooled into several African regions (see table included in supplemental file S1). We used the DnaSP 5.10 program (Librado and Rozas 2009) to estimate diversity indices (except ϑ_K) and Arlequin 3.11 software (Excoffier et al. 2005) to obtain ϑ_K values and pairwise F_{ST} genetic distances. We carried out admixture analyses assuming that the mtDNA variation accumulated in the different Ugandan populations comes from one of the six main African regions: North, West-Central, Eastern, Southern, Southeast and Southwest Africa.

We also generated whole-mtDNA genome sequences in 15 samples belonging to the new haplogroup L3i1 (Table S3). These sequences were combined with previously published ones into a total data set of 27 L3i whole-mtDNAs. The new sequences are deposited in GenBank: KP229441–KP229455. Within L3i, we estimated ages of clades using a time-dependent clock, incorporating a correction for purifying selection (Soares et al. 2009) using both the ρ statistic (Forster et al. 1996) and maximum likelihood (ML) with PAML 3.13 (Yang 1997) (Table S4).

To investigate the overall genomic ancestry in two samples belonging to haplogroups originating outside Africa (HV1b1 and T1a), we used 46 autosomal ancestry informative markers (AIMs), previously described as efficiently inferring proportions of African, European, East Asian and Native American ancestries (Pereira et al. 2012).

We performed principal component analysis (PCA) of the indigenous (LxMN) mtDNA haplogroup frequencies using XLSTAT for the Ugandan samples distinguished by language (Eastern Nilotic ENU, Western Nilotic WNU and Bantu speakers BUG), in the context of population data from across Africa.

We examined in detail only the first three components (PC1, PC2, and PC3) from the analysis and we visualized the output using XLSTAT. PC3 did not add to the interpretation and is not discussed. We checked F_{ST} values between populations and also confirmed that the results were broadly similar to those obtained using multidimensional scaling [see Silva et al. (2015), Fig. 3]. We included only the following haplogroups: L0a, L0b, L0d, L0f, L0k, L1b, L1c, L2a, L2b, L2c, L2d, L2e, L3a, L3b, L3c, L3d, L3e, L3f, L3h, L3i, L3x, L4a, L4b, L5a, L5b and L6, excluding unclassified and paraphyletic lineages. We also excluded populations of sample size below 40 and also very recently admixed or highly diverged/drifted populations of the south and from the rainforest zone. The samples included in PCA are listed in table included in supplemental file S1.

We displayed genetic relationships between HVS-I haplotypes from the present study and 9034 from the literature (see supplemental file S1 for data source), within each haplogroup, by means of phylogenetic networks generated by post-processing the output of the reduced-median algorithm with the median-joining algorithm (Bandelt et al. 1999; Bandelt et al. 1995) using Network 4.6.1.1 software (http://www.fluxus-engineering.com), resolving them guided by the whole-mtDNA genome phylogeny in PhyloTree Build 16, 19 Feb 2014 (van Oven and Kayser 2009). Even using this approach, it was difficult to match the known phylogeny exactly in every instance of branching, so that not every case of deep branching could be correctly represented. We estimated the age of clades of interest using the ρ statistic with Network, using a mutation rate for HVS-I of one mutation in 16,667 years

(Soares et al. 2009). For the network analyses, we grouped the samples on a regional basis as for the PCA, but were able to also include additional populations with lower sample size.

Full details on samples and methods used are available in supplemental file S1.

Results

Diversity of Ugandan mtDNAs

Using the 16090–16365 sequence range of the HVS-I region we found a total of 228 different mtDNA haplotypes in our sample of 409 individuals from Uganda (Table S2). The haplogroup distribution is shown in Figure 1. The overall haplotype diversity was high (0.977±0.002). The genetic diversity found in Uganda was higher than that observed in East, North, West-Central (Bantu and Pygmies) and South, but similar to values obtained for the most diverse regions in sub-Saharan Africa.

When considering the samples from Bantu and the Eastern and Western Nilotic speakers separately, we observed no significant difference in the diversity values between these linguistic groups (Table S2). Since their arrival in the Great Lakes region, Bantu-speaking groups have been concentrated in the South, with Western Nilotic speakers inhabiting primarily the Northwest part of the West Nile region. Nevertheless, due to competition for land in which to practice agriculture, it is thought that the Bantu speakers interacted more with Western than with the Eastern Nilotic speakers, who were concentrated in the Northeast and many of whom retained a more pastoral economy (Maxon 2009; Newman 1995). Political disputes during the seventeenth and eighteenth centuries contributed to the complex interaction between these Bantu and Western Nilotic communities, increasing the opportunities for intermarriage (Maxon 1994) and consequently a higher diversity for these groups might be expected. Nevertheless, we observe the highest values of nucleotide diversity and average number of pairwise differences in the sample from Eastern Nilotic speakers. The distinct pastoralist lifestyle of the Eastern Nilotic speakers – in particular, the Karamojong, included in the present study – led to less contact with other ethnic groups in Uganda, although the extent to which the expansions may have assimilated other peoples is not known. The Nyangatom from Ethiopia who, like the Karamojong, also belong to the Teso-Turkana branch of the Eastern Nilotic languages, showed similarly high diversity, although with slightly lower values than that from the Karamojong (Poloni et al. 2009).

Principal component analysis of Ugandan mtDNAs

PC1 (21.7% of the variance) portrays a broadly East–West axis, and PC1 and PC2 together (34.7% in total) sequester the populations into four distinct quadrants that make excellent historical and geographical sense (Fig. 2).

The upper-left-hand **"Eastern" quadrant**, is defined by elevated levels of haplogroups L0b, L3c, L3i, L3x, L4a, L5a, L5c and L6. It includes populations from Eastern Africa, primarily the Horn (in medium pink) and Sudan (in dark pink), and includes as well all of the Eastern Nilotic populations (in orange), with those of Ethiopia clustering with other Ethiopian populations (including Afroasiatic speakers) and those of Kenya and Uganda clustering more closely with Somalia and (to a lesser extent) Sudan. Sudan is somewhat distinct from the Horn populations, very likely because it experienced heavy immigration in the mid-Holocene from West-Central Africa, mimicking the later impact of the Bantu dispersals further south; when excluding the inferred immigrant L2a and L0a lineages it indeed clusters closely with Horn populations (Silva et al. 2015). The Eastern Nilotic-speaking Ugandans are quite distinct from the other Ugandan groups, but along with Eastern Nilotic speakers from Kenya they fall very close to other Eastern Africans. They are, however, shifted towards the East African Bantu cluster, which may be due to the heavy gene flow with Ugandan Bantu and Western Nilotic speakers (see below).

The upper-right-hand **"West-Central/North" quadrant** is defined by elevated levels of L1b, L2a-c, L2e, and L3f. It includes all populations classified as West-Central (in grey) or North African (in green), except for Egyptians (EGY) and the Mbuti (PMB), who are shifted to the east, a slight spill-over of the Nigerian point (NIG) into the lower-left-hand "Bantu" quadrant, and Western Nilotic speakers from Uganda (WNU, in yellow). The sub-Saharan component of most North Africans probably crossed the Sahara from West-Central Africa during the pluvial phase of the early to mid-Holocene, explaining their close links to West-Central Africans now (see also the comment on Sudan, above). Egypt, unlike other North African regions, probably received its sub-Saharan component from Eastern rather than West-Central Africa. The Mbuti live in the Ituri rainforest in the northeast of the Democratic

Republic of Congo, very close to Uganda, and many speak Sudanic languages, so a more easterly position makes very good sense (and they are also highly drifted due to small population size).

Thus the overall clustering makes strong historical sense, and implies a partly West-Central African source for Western Nilotic speakers from Uganda, although caution is warranted here because of the small samples size (47). Since they are shifted towards the Eastern quadrant compared with other West-Central African populations, a substantial Eastern component, perhaps from Eastern Nilotic speakers, (or even suggesting a partial origin in Sudan) may also be suggested.

The lower-right-hand **"Bantu" quadrant** is defined by elevated levels of L1c and L3e. It comprises almost entirely of Bantu-speaking populations from West-Central/Southwest (dark blue) and Southeast (indigo) Africa, which all cluster together fairly tightly, implying a largely common source. Those from West-Central/Southwest Africa (Cameroon, Gabon, Cabinda and Angola) are, except for Angola, closer to the West-Central/North quadrant, whilst those from Southeast Africa are shifted over slightly towards, towards the fourth quadrant, implying a minor contribution from East Africa. The only exceptions are the Kenyan Mijikenda (BKM), who have been previously noted as relatively unassimilated compared to other East African Bantu-speaking groups (Batai et al. 2013), and even these are shifted somewhat towards the eastern pole.

The fourth, lower-left-hand quadrant, which we have described as the **"melting-pot" quadrant**, is defined by elevated levels of L0a, L0f, L3a, L3h and L4b. It comprises a diverse set of almost entirely East African populations. We have discussed the Mbuti and Egyptians already, above. Now we see Kenyan and Tanzanian Afroasiatic Cushitic speakers (KAA and TZH, respectively, in pale pink) at the lower-left-hand extreme (notably at the opposite pole in PC2 from Afroasiatic speakers from Ethiopia, ETA/ETH), Tanzanian click-speaking Sandawe (TZS, in purple) and Southern Nilotic speakers (SNT, in red) towards the centre of the quadrant and most East African Bantu-speaking groups (in turquoise), including from Rwanda as well as Uganda, Tanzania and Kenya (Taita), in the upper-right-hand part of the quadrant. Clustering with these are the Kenyan urban (Nairobi) sample (KUR, in hatched pink), the Tanzanian click-speaking Hadza (TZH, in purple) and the Kenyan Western Nilotic-speaking Luo (WNL, in yellow). The Hadza are particularly affected by drift, but their association with the Bantu cluster may in part reflect the significant fraction (almost a quarter) of L2a in the Hadza (almost absent from the Sandawe), as well as L3b and L3e, some or all of which may be the result of gene flow with Bantu speakers, which is also suggested by genome-wide analyses (Tishkoff et al. 2009), but it may also to some extent reflect the substrate gene pool assimilated by the Bantu speakers in East Africa (*e.g.* L0f, L3a, L3h, L3i, L4, L5), which may similarly be implied by the autosomal evidence (Tishkoff et al. 2009).

The plot therefore suggests that the East African Bantu-speaking populations, apart from the Kenyan Mijikenda discussed above, have a mosaic ancestry that includes components from both the indigenous populations and incoming Bantu speakers from West-Central Africa. The Ugandan Bantu speakers appear to have a larger fraction of West-Central African ancestry than either the Kenyan Taita or the Tanzanian sample, suggesting increasing levels of assimilation *en route* (as in classical models of demic diffusion). Both the Nairobi sample and the Western Nilotic-speaking Luo from Kenya may have a predominantly Bantu-speaking ancestry, and the Luo appear to have a distinct ancestry from the Western Nilotic speakers of Uganda. The position of the Tanzanian Southern Nilotic speakers implies a largely indigenous origin or very extensive gene flow with neighbouring populations, in line with the autosomal picture (Tishkoff et al. 2009).

Summarising the results from the first two components, the Eastern and Western Nilotic and Bantu speakers from Uganda are very distinct, with the Eastern Nilotic speakers (ENU) clustering more closely with East African, Horn and Sudan populations, Western Nilotic speakers (WNU) with West-Central Africans, and Bantu speakers (BUG) with other Bantu speakers from East Africa. Nevertheless different relationships emerge with each major component, and more detailed analyses are needed in order to tease them apart.

Admixture analysis

We also carried out admixture analyses in order to estimate the contribution of six different African regions to the four linguistic groups from Uganda, assuming they could broadly be considered as potential source populations (Table 1). The analyses were based on haplotype sharing between the different regions, performed in three slightly different ways: considering perfect matches (P0), and one- or two-mutational step neighbouring haplotypes

(P1or P2), between the source and the case-study populations. As shown in Table 1, the three estimates yielded similar results.

The results indicate that West-Central Africa is the main contributor to both the Bantu and Western Nilotic populations, closely followed by Eastern Africa. In contrast, the Eastern Nilotic speakers showed the opposite pattern, although still with a substantial West-Central African component. North Africa also appears to have contributed to the Ugandan populations with a substantial proportion in all four groups (~10%), although it has to be considered that these analyses assume no gene flow in the reverse direction, which might better explain this particular pattern. Estimates for the Central Sudanic-speaking population are less reliable due to the small sample size. Nevertheless, the three admixture estimates consistently indicate a significant contribution from West-Central Africa, which would fit well with the present distribution of the Central Sudanic languages. As a whole, it is worth noting the contribution of a major Central African substrate to the ancestry of Ugandan populations.

Phylogeography of Ugandan lineages

Within Africa, a number of mtDNA lineages have been associated with various stages of the Bantu dispersals, including subsets of L0a (Pereira et al. 2001; Watson et al. 1997), L1c (Batini et al. 2007; Salas et al. 2002), L2a (Pereira et al. 2001; Salas et al. 2002), L3b (Salas et al. 2002; Soares et al. 2012; Watson et al. 1997) and L3e (Bandelt et al. 2001; Pereira et al. 2001; Plaza et al. 2004; Salas et al. 2002; Soares et al. 2012). Many of these are inferred to have spread from West-Central Africa, although some, e.g. L0a, were considered likely assimilated in Central or East Africa. The distribution of haplogroups L0f, L3f, L5a, L5b, L6 and L4b2, as well as L0a, appears to be centred on Eastern African populations, with L3i1 seen mainly in Eastern Nilotic speakers (Castri et al. 2009; Černý et al. 2007; Coudray et al. 2009; Kivisild et al. 2004; Poloni et al. 2009; Salas et al. 2002). L0d and L0k are characteristic of Khoisan-speaking populations (Behar et al. 2008; Černý et al. 2007; Plaza et al. 2004; Quintana-Murci et al. 2010; Salas et al. 2002), and within the largely Central African haplogroup L1c, at least L1c1a was considered to have an origin amongst Western Pygmy ancestors (Batini et al. 2007; Quintana-Murci et al. 2010).

The 409 mtDNAs sequenced during the present study were classified into haplogroups (Table S1). 98% belong to (LxMN) haplogroups. The haplogroup networks (Fig. 3

and supplemental Figs. S1-S11) have two main qualities: they both show a huge amount of gene flow between populations and regions of Africa, and yet at the same time they show strikingly different patterns of lineage distributions, indicating that they can be highly informative on the history of the continent. A fuller overview of the haplogroups is available as supplemental material (see supplemental file S2).

In Uganda, individuals belonging to **haplogroups L0a and L0f** (Figs. S1 and S2) are mainly Bantu or (much more commonly) Eastern Nilotic speakers, with none in Western Nilotic speakers, despite their being common in the Kenyan Western Nilotic Luo (Castri et al. 2008). The distribution of Ugandan L0a lineages in the network suggests that many of the Eastern Nilotic L0a lineages most likely arose directly within pre-Bantu East Africa, whereas the Bantu lineages originated mostly as a part of the Bantu dispersals from West-Central Africa but with a substantial minority assimilated within East Africa – seen most clearly in L0f, but also in L0a1d. This points to a distinct origin for the Western Nilotic speakers, who lack this major signal of East African maternal lineages. A diverse set of L0a lineages is, however, also present at extremely high frequency (>50%) in the small sample of Tanzanian Southern Nilotic speakers, the majority of which match Eastern Nilotic speakers from East Africa, suggesting a common East African genetic stratum for both of these two language groups.

Haplogroup L1c was concentrated especially in Nilotic speakers in Uganda, particularly the Western group (see supplemental Fig. S3). The distribution suggests that the L1c lineages were brought to the Great Lakes from West-Central Africa and then assimilated into Nilotic speakers.

The largest **haplogroup L2a** subclade found in Uganda is L2a1 (supplemental Fig. S4). The few Eastern Nilotic lineages within L2a1 are not obviously of Bantu origin, and may have been in Eastern Africa prior to the Bantu arrivals. We note that although L2a dispersed from West-Central Africa with the Bantu, it is also thought to have dispersed earlier from the same region, in the early to mid-Holocene, via the “Green Sahara” into Sudan and, to a lesser extent, the Horn of Africa (Kuper and Kröpelin 2006; Silva et al. 2015) (see also the PCA results, above).

Most of the Ugandan L2a variation is concentrated within a single subclade, L2a2’3’4, which dates to ~30 ka (Silva et al. 2015) and is much more prevalent in Central and East

Africa than other L2 subclades. Intriguingly, many of the Ugandan Eastern and Western Nilotic L2a lineages within the subclade L2a2'3 either match or are closely related to Sudanese – and indeed many of the Ugandan Bantu lineages, too. Whether this might be an indication of a Sudanese source for a fraction of Ugandan diversity merits further investigation, but a source somewhere in Eastern Africa seems likely. L2a4, dating to ~24 ka (Silva et al. 2015) appears likely indigenous to East Africa itself, and L2a5, which dates to ~46 ka, is also Eastern African and includes several Ugandan Bantu and Nilotic speaking individuals, including one lineage that has also reached Southern Nilotic speakers of Tanzania. Corroborating this interpretation, most of the few sampled Central Sudanic speakers from Uganda, whose languages are assumed to predate the appearance of Nilotic and Bantu languages in the region, also fall within these L2a subclades, again suggesting an earlier expansion.

As with L2a, the predominantly West-Central African **haplogroups L2b, L2c, L2d and L2e** (Fig. S5) are substantially more frequent in the Western than the Eastern Nilotic groups of Uganda, with the Bantu roughly in between for L2a, but with a frequency of other L2 lineages similar to the Eastern Nilotic speakers. Probably all dispersed from West-Central Africa during the Holocene, but whilst some suggest Bantu dispersals (e.g. in L2d), many of them appear to have dispersed independently (e.g. in L2e).

The preponderance of **haplogroup L3b** in Ugandan Bantu speakers and their distribution amongst the three Ugandan groups (Fig. S6) suggests an introduction to the Nilotic speakers from the incoming Bantu arrivals from West-Central Africa, with at least three founder haplotypes in Nilotic speakers. For the much rarer **haplogroup L3d** (Fig. S7) found mainly in Ugandan Bantu and Western Nilotic speakers, some lineages in Uganda are likely due to Bantu dispersals (e.g. in L3d3a and L3d1a1'2), but some lineages found in Nilotic speakers fall within apparently non-Bantu West-Central African clusters, which may, as with L2e, be due to distinct, non-Bantu-mediated dispersals from West-Central Africa.

Haplogroup L3e (Fig. S8) is twice as frequent in the Bantu that in the Nilotic speakers from Uganda, and again can be interpreted as mainly due to the Bantu expansion from West-Central Africa, and gene flow into Nilotic speakers, with some signs of very recent gene flow from the Eastern to Western Nilotic-speaking Ugandans.

Haplogroup L3f (Fig. S9) is rare in the Ugandan samples and almost entirely restricted to Eastern Nilotic speakers. Curiously, the Ugandan lineages do not fall into the Eastern African subclades but instead into L3f1b and L3f3, with a hint of a possible link to Sudan *via* the Sahara that would merit further investigation (Černý et al. 2009).

There is also a collection of rare Eastern African lineages in the Ugandan samples within haplogroup L3, mostly restricted to Eastern Africa, which are poorly defined within HVS-I and not straightforward to portray accurately in a phylogenetic network (Fig. 3). Even so, with the aid of the whole-mtDNA phylogeny and the available control-region information, we can readily draw some general conclusions from their distinctive distributions.

They include **hapologroup L3a, haplogroup L3x** and several more unclassified L3* lineages, most of which have an ancestry in East Africa (Fig. 3). Several subclades include diverse lineages within Eastern Nilotic-speaking Ugandans, with sporadic gene flow into Ugandan Western Nilotic or Bantu-speaking groups. Of particular interest, however, are haplogroups L3h and L3i (Fig. 3), both of which highly informative, albeit pointing in different directions. Although rare, **haplogroup L3h** is present across Eastern Africa (Kivisild et al. 2004; Poloni et al. 2009; Soares et al. 2012; Tishkoff et al. 2007). It dates to ~55.1 ka (Soares et al. 2014) with the two basal subclades L3h1 and L3h2 dating to ~47 ka (Behar et al. 2012) and L3h2 to ~31 ka (from our HVS-I network). Intriguingly, all of the major identifiable subclades of L3h are found at high diversity in Sudan, as well as across the rest of Eastern Africa.

L3h2 is more frequent in the Horn of Africa, but also includes a number of Kenyan Eastern Nilotic and Western Nilotic speakers. There is low diversity in both the Horn and Nilotic lineages, suggesting a possible recent movement from Sudan into the Horn and East Africa (albeit with few data). All of the Ugandan L3h lineages fall into L3h1a1 and, in contrast with the Kenyans, display very high diversity. Nevertheless, the diversity is again also high in Sudan – and there are several direct matches – suggesting that a possible source there would be worth pursuing. One L3h1 haplotype that matches a Ugandan Eastern Nilotic lineage is seen in the Western Nilotic speakers and another is seen in a Ugandan Bantu speaker, likely due to assimilation within East Africa from Eastern Nilotic speakers, who are an important reservoir for L3h1a diversity. L3h1b is atypical within L3h, as it includes a number of West-Central African lineages, including an entire subclade, L3h1b2, dating to ~13 ka (Behar et al. 2012), and three major East African Bantu haplotypes that most likely derive from them; one Ugandan Bantu lineage appears to derive from this source, quite distinct from the Nilotic lineages, whereas others suggest assimilation within East Africa. However, L3h1b is also intriguing for including a high frequency of lineages belonging to Tanzanian Southern Nilotic speakers, and several diverged lineages from Sudan. This is the only clear indication we have in the mtDNA that some shallow lineages in Southern Nilotic as well as Eastern Nilotic speakers may have spread from Sudan in the relatively recent past.

Haplogroup L3i1 lineages (Fig. 3; L3i* lineages cannot be identified from HVS-I screening) have been previously described in Ethiopia (~3%), two-thirds of them in Eastern Nilotic and the remainder in Afroasiatic speakers (Kivisild et al. 2004; Poloni et al. 2009). Most of the L3i1 lineages identified here belong to a distinctive branch of L3i, not labelled in PhyloTree, which we here refer to as L3i1c. This subclade is virtually restricted to Ugandan Eastern Nilotic speakers (where it occurs at a rate of ~7%), despite harbouring substantial diversity. It is entirely absent from the Ugandan Bantu samples, although it occurs in several Taita from kenya, but not in other Kenyan Bantu speakers (Batai et al. 2013). The remainder of L3i1, mostly single individuals, is seen Kenya (a Bantu Kikuyu), and non-Bantu speakers in Ethiopia, Sudan, Egypt and also far to the West in Mauritania. There is also a single Eastern Nilotic speaker within the Eastern African L3i2 sister subclade, which is distributed mainly across Kenya (including a very few Kikuyu Bantu speakers – again, the only ones found carrying the subclade) and the Horn, with a few in North Africa and one in Sudan.

Given the paucity of data for this lineage, we sequenced the whole-mtDNA genome of one L3i1a and 14 L3i1c samples from Uganda, in order to reconstruct a phylogenetic tree of L3i as a whole (Fig. 4). This allowed us to date L3i1c using ML to ~14.2 ka, L3i1 as a whole to ~27 ka, L3i2 to ~14.4 ka, and L3i overall to ~43 ka, the latter in close agreement with Behar et al. (2012) (Table S4). This indicates that the Ugandan subclade L3i1c dates to the Late Glacial period.

Within **haplogroup L4**, L4b2 was the only subclade found in our sample (~10%) (Fig. S10) reaching the highest frequency and by far the highest diversity amongst Eastern Nilotic speakers. Haplotype sharing and diversity patterns point to an East African source for mtDNAs found in Eastern Nilotic speakers, assimilated by the Ugandan Bantu within East

Africa. The rare lineages found in Western Nilotic speakers can all be attributed to recent gene flow with the Eastern Nilotic speakers. Similar to other haplogroups carried at high frequencies by the Tanzanian Southern Nilotic-speaking sample, there is a strong degree of haplotype sharing with Eastern Nilotic speakers from East Africa, including Uganda. Overall a source in East Africa or the Horn for the Eastern Nilotic lineages is indicated; the very minor links to Sudan are more likely the result of gene flow in the reverse direction.

In Uganda, **haplogroup L5** encompassed both L5a and L5b lineages (Fig. S11). Within Uganda, L5b is present exclusively in the Nilotic speakers, and L5a again seen mainly in Eastern Nilotic speakers, with a single haplotype in Bantu speakers shared with non-Bantu Eastern Africans. This pattern again suggests a source for Eastern Nilotic speakers within Eastern Africa, and minor gene flow towards Bantu speaking groups and also towards West-Central Africa. Although possibly an effect of small sample size, Eastern Nilotic speakers in Uganda and, especially, Horn populations, appear more diverse than other East African L5 lineages – a possible hint that a small fraction of the maternal ancestry of Eastern Nilotic speakers might trace to the East or North, in approximately the direction from which the languages are thought to have arisen. Furthermore, there are close connections between Eastern Nilotic speakers and Sudanese lineages in both L5a and L5b which would again merit further study.

Phylogeography of "Eurasian" lineages in Uganda

There are a number of Eurasian mtDNA lineages (M1a, N1a1, HV1b1, R0a1a, J1d1 and T1a) at low frequencies in Ugandans (both Eastern Nilotic and Bantu speakers) of West Asian and possibly even European provenance (Fig. 1; supplemental Table S1). This is reflected in a substantial Eurasian autosomal component in Eastern Africa, which is also evident to some extent in Kenya (Pagani et al. 2012). Due to deep ancestry and the distribution of these lineages (Abu-Amero et al. 2008; Brakez et al. 2001; Černý et al. 2008; Coudray et al. 2009; Fernandes et al. 2012; Kivisild et al. 2004; Olivieri et al. 2006; Rhouda et al. 2009; Richards et al. 2000; Richards et al. 2003), none of the West Eurasian mtDNAs found in Uganda has a likely European source, for example resulting from the heavy European colonial involvement since the 1870s (Maxon 2009). One possibility is a Bronze-Age dispersal from the Near East accompanying the spread of Semitic languages (Kitchen et al. 2009), consistent with an inferred Levantine (rather than Arabian) source (Pagani et al. 2012). However, the T1a lineage is a possible candidate; and it is also possible that the Arabian lineages in Uganda, such as HV1b1, might only have arrived very recently in the Great Lakes region.

We therefore used 46 autosomal AIMs to explore further the ancestry of the individuals carrying HV1b1 and T1a, in order to test whether or not their presence could be due to modern admixture. STRUCTURE and PCA showed that all Ugandan samples tested have a clear sub-Saharan African ancestry, with a close relationship with the HGDP–CEPH sub-Saharan individuals used as training set (see supplemental file S3 for detailed information on the results obtained by different analyses). These results show that the HV1b1 and T1a mtDNA sequences found in Uganda are not the result of recent maternal European influx within the last two generations. This is in accordance with the results from a previous study of paternal composition of the same Eastern Nilotic population showing the presence of a single non-African chromosome carrying the M70 mutation (Gomes et al. 2010), which was also found in other Eastern African populations. The West Eurasian lineages in Uganda were therefore most likely the result of more ancient movements of people to Eastern Africa from North Africa and/or Arabia, where HV1b1 is generally found (Richards et al. 2003).

Discussion and Conclusions

This study shows that not only does the Great Lakes region of Uganda harbour enormous genetic diversity but that, despite significant levels of recent gene flow, the various linguistic groups are also strongly differentiated from each other. The results point to a mosaic of diversified groups that have experienced an extremely complex history in which dispersal, gene flow and acculturation went hand in hand.

The two Nilotic-speaking groups are highly differentiated and show few signatures of common ancestry within the last few millennia, although there is clear evidence for gene flow between them. This appears mainly to be from the Eastern to the Western group, although some caution is needed due to the smaller sample size of Western Nilotic speakers and the reduced likelihood of sampling rarer lineages. Both groups retain very high levels of genetic diversity, which may be accounted for by assimilation of lineages from both local populations and Bantu-speaking groups. Nevertheless, the high diversity values observed may be also the result of geographic and/or ethnic structure inside of which linguistic group.

Could any of these lineages be markers for dispersals associated with the spread of Eastern Nilotic speakers? Nilotic-speaking populations are thought to have emerged in Southern Sudan by the beginning of the first millennium BC (Ehret 1998), with Eastern Nilotic speakers arriving in Northeast Uganda via Kenya within the last few hundred years (Maxon 2009; Newman 1995). The potential links to Sudan in L3h1a, L5a and, more tentatively, in L2a2’3, suggest that a Sudanese source for a small minority of the lineages in Ugandan Eastern Nilotic speakers remains a strong possibility. Further detailed study of these lineages at the whole-mtDNA level may prove extremely illuminating in teasing out details of these dispersals that are currently inaccessible to non-genealogical whole-genome studies.

On the other hand, both the HVS-I network and the whole-genome tree for L3i clearly show that, although an origin in the Horn is certainly quite possible, the Ugandan subclade, L3i1c, itself dates to the Late Glacial period. In this case, there is no signal of a genetic trail from further north.

In principle, there are at least two possible explanations for this pattern. Firstly, the entire cluster may have been brought from the Horn of Africa recently, but if this were the case then no genetic trail has survived: no lineages in Kenya, Ethiopia, Somalia or Sudan that cannot be accounted for by very recent gene flow. It seems unlikely that a population should move *en masse* without leaving any trace in the lands from whence it came. But even if the entire community was indeed involved in such a mass migration, the timescale for the Ugandan L3i1c subclade is ~14 ka: previous generations would certainly have experienced gene flow with neighbouring populations, leaving related lineages to evolve across the region. Thus this explanation does not seem plausible.

A second possibility, then, is that Eastern Nilotic language speakers, or at least a population that became Eastern Nilotic speaker, have an at least partial ancient, Late Glacial ancestry within Uganda itself. If this were the case, they must have been present before the arrivals of Bantu and Western Nilotic speakers, since neither of the other two language groups carry this haplogroup to a significant extent. This seems a much more straightforward phylogeographic interpretation.

The network (Fig. 3) shows that the Ethiopian and Kenyan samples carry HVS-I haplotypes shared with the Ugandans. Ugandan samples are much more diverse, suggesting assimilation of ancient indigenous lineages by the Eastern Nilotic speakers in Uganda and a more recent dispersal in the opposite direction, north-eastwards from Uganda into Kenya and as far as Ethiopia. L3i1's sister clade L3i2 also appears to have a Late Glacial source in Eastern Africa, possibly the Horn, judging from the HVS-I network, dating to ~17.2 ka with whole-mtDNAs (mainly from the Horn; 13.7 ± 4.9 ka with the larger and geographically much more diverse number of HVS-I sequences).

Thus, L3i1c points to recent long-distance gene flow between Eastern Nilotic-speaking populations in the opposite direction to that of the historically attested dispersals into Uganda from Kenya. In addition, within Uganda, one L3i1c haplotype seems to have been recently introduced into Western Nilotic speakers, which otherwise lack L3i. The few samples in East African Bantu speakers similarly suggest assimilation from indigenous East Africans. The absence of L3i1 in Southeast African Bantu groups (e.g., Salas et al. 2002) indicates that these lineages were not assimilated during the Bantu expansion, but more recently.

.The Bantu speakers in Uganda display a high mtDNA diversity that cannot be ascribed simply to either an ancient West-Central or East African ancestry, but has been enriched by distinct ancestral genetic substrates from across the continent, which in turn bequeathed a substantial part of this ancestry to the indigenous populations of the Great Lakes region.

More broadly, this study emphasises that, even in the genomic age, the analysis of the distribution of mtDNA control-region sequences, provided that it is underpinned by a detailed understanding of the whole-mtDNA phylogeny, remains a valuable first step towards reconstructing the demographic history of a genetically uncharacterised population.

Acknowledgments

This study was supported by Fundação para a Ciência e Tecnologia (FCT) and co-financed by the European Social Fund (Human Potential Thematic Operational Programme) through grants SFRH/BPD/76207/2011 (VG) and SFRH/BPD/81986/2011 (RP). PS is supported by FCT, European Social Fund, Programa Operacional Potencial Humano and the FCT Investigator Programme (IF/01641/2013) and acknowledges FCT/MEC for support to CBMA through Portuguese funds (PIDDAC) - PEst-OE/BIA/UI4050/2014. AS is supported by "Ministerio de Ciencia e Innovación" (SAF2011-26983), the Plan Galego IDT (EM 2012/045) and the grant from the Sistema Universitario Gallego- Modalidad REDES (2012-PG226) from the Xunta de Galicia. IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, the Portuguese Foundation for Science and Technology. This work is funded by FEDER funds through the Operational Programme for Competitiveness Factors - COMPETE and National Funds through the FCT - Foundation for Science and Technology, under the project "PEst-C/SAU/LA0003/2013". NORTE-07-0162-FEDER-00018 - Contributos para o reforço da capacidade do IPATIMUP enquanto actor do sistema regional de inovação" and NORTE-07-0162-FEDER-000067 - Reforço e consolidação da capacidade infraestrutural do IPATIMUP para o sistema regional de inovação", both supported by Programa Operacional Regional do Norte (ON.2 – O Novo Norte), through FEDER funds under the Quadro de Referência Estratégico Nacional (QREN). We would like to acknowledge the contribution of Magdalen Awor, Fr. Germano Serra and Iva Gomes for collecting the Karamoja samples. We thank Eric Bridgeland for collecting the Kabale samples, Dr Narcis Kabatereine (Vector Control Division, Uganda Ministry of Health) for the Piida samples, and Abigail Enaburekhan, Sadie Anderson-Mann, Rebecca Cole and Steven Groom for preliminary work on the Kabale and Piida sequences.

Compliance with ethical standards

Conflict of interest: The authors declare that they have no conflict of interest.

Research involving human participants: All procedures performed in studies involving human participants were approved by the University of Leeds, Faculty of Biological Sciences Ethics Committee, and that of the University of Huddersfield, School of Applied Sciences.
Informed consent: Informed consent was obtained from all individuals participants included in the study.

References

Abu-Amero KK, Larruga JM, Cabrera VM, González AM (2008) Mitochondrial DNA structure in the Arabian Peninsula. BMC Evol Biol 8:45

Ambrose SE (1982) Archaeology and linguistic reconstructions of history in East Africa. In: Ehret C, Posnansky M (eds) The archaeological and linguistic reconstruction of African history. University of California Press, Berkeley, pp 104-157

Atkinson QD, Gray RD, Drummond AJ (2008) mtDNA variation predicts population size in humans and reveals a major southern asian chapter in human prehistory. Molecular Biology and Evolution 25: 468-474. doi: 10.1093/molbev/msm277

Balter M (2011) Was North Africa the launch pad for modern human migrations? Science 331: 20-23. doi: 10.1126/science.331.6013.20

Bandelt H-J, Alves-Silva J, Guimarães PEM, Santos MS, Brehm A, Pereira L, Coppa A, Larruga JM, Rengo C, Scozzari R, Torroni A, Prata MJ, Amorim A, Prado VF, Pena SDJ (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet 65: 549-563.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37-48.

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141: 743-53.

Batai K, Babrowski KB, Arroyo JP, Kusimba CM, Williams SR (2013) Mitochondrial DNA diversity in two ethnic groups in Southeastern Kenya: Perspectives from the northeastern periphery of the bantu expansion. American Journal of Physical Anthropology 150: 482-491. doi: 10.1002/ajpa.22227

Batini C, Coia V, Battaggia C, Rocha J, Pilkington MM, Spedini G, Comas D, Destro-Bisol G, Calafell F (2007) Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa. Molecular Phylogenetics and Evolution 43: 635-644.

Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, Kivisild T, Torroni A, Villems R (2012) A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90: 675-84. doi: 10.1016/j.ajhg.2012.03.002

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S (2008) The dawn of human matrilineal diversity. The American Journal of Human Genetics 82: 1130-1140.

Beleza S, Gusmao L, Amorim A, Carracedo A, Salas A (2005) The genetic legacy of western Bantu migrations. Hum Genet 117: 366-75. doi: 10.1007/s00439-005-1290-3

Bender LM (2000) Nilo-Saharan in B. Heine, and N. D., eds. African languages: An introduction. Cambridge University Press, Cambridge, pp 43-73

Boattini A, Castrì L, Sarno S, Useli A, Cioffi M, Sazzini M, Garagnani P, De Fanti S, Pettener D, Luiselli D (2013) mtDNA variation in East Africa unravels the history of Afro-Asiatic groups. American Journal of Physical Anthropology 150: 375-385. doi: 10.1002/ajpa.22212

Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, Calafell F (2001) Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. Annals of Human Biology 28: 295-307.

Campbell MC, Tishkoff SA (2010) The evolution of human genetic and phenotypic variation in Africa. Current Biology 20: R166-R173.

Castri L, Garagnani P, Useli A, Pettener D, Luiselli D (2008) Kenyan crossroads: migration and gene flow in six ethnic groups from Eastern Africa. Journal of Anthropological Sciences 86: 189-192.

Castri 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: 302-11. doi: 10.1002/ajpa.21070

Černý V, Fernandes V, Costa MD, Hajek M, Mulligan CJ, Pereira L (2009) Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. BMC Evol Biol 9: 63. doi: 10.1186/1471-2148-9-63

Černý V, Mulligan CJ, Rídl J, Zaloudkova M, Edens CM, Hájek M, Pereira L (2008) Regional differences in the distribution of the Sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. Am J Phys Anthropol 136: 128-137.

Černý V, Salas A, Hajek M, Zaloudkova M, Brdicka R (2007) A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. Ann Hum Genet 71: 433-52. doi: 10.1111/j.1469-1809.2006.00339.x

Coia V, Destro-Bisol G, Verginelli F, Battaggia C, Boschi I, Cruciani F, Spedini G, Comas D, Calafell F (2005) Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. Am J Phys Anthropol 128: 678-81. doi: 10.1002/ajpa.20138

Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A, Dugoujon JM (2009) The complex and diversified mitochondrial gene pool of Berber populations. Ann Hum Genet 73: 196-214. doi: 10.1111/j.1469-1809.2008.00493

Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, Scozzari R (2011) A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. Am J Hum Genet 88: 814-8. doi: 10.1016/j.ajhg.2011.05.002

David N (1982) Prehistory and historical linguistics in central Africa: Points of contact. In: Ehret C, Posnansky M (eds) The archaeological and linguistic reconstruction of African history. University of California Press, Berkeley, pp 78-103

Ehret C (1998) An african classical age. James Currey, Oxford

Ehret C (2001) A historical-comparative reconstruction of Nilo-Saharan. Rüdiger Köppe Verlag, Cologne

Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: An integrated software package for population genetics data analysis. Evolutionary bioinformatics online 1: 47-50.

Fernandes V, Alshamali F, Alves M, Costa Marta D, Pereira Joana B, Silva Nuno M, Cherni L, Harich N, Cerny V, Soares P, Richards Martin B, Pereira L (2012) The arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. The American Journal of Human Genetics 90: 347-355. doi: 10.1016/j.ajhg.2011.12.010

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of native American mtDNA variation: A reappraisal. American Journal of Human Genetics 59: 935-945.

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

Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA (2007) Whole-mtDNA genome sequence analysis of ancient African lineages. Mol Biol Evol 24: 757-68. doi: 10.1093/molbev/msl209

Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME (2008) Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history. American Journal of Physical Anthropology 137: 316-323.

Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodriguez-Botigue L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK, Norman PJ, Parham P, Bustamante CD, Mountain JL, Feldman MW (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci U S A 108: 5154-62.

Kitchen A, Ehret C, Assefa S, Mulligan CJ (2009) Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East

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. American Journal of human genetics 75: 752-770.

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sjantila A, Paabo S, Stoneking M (1999) mtDNA analysis of Nile river valley populations: A genetic corridor or a barrier to migration? American Journal of human genetics 64: 1166-1176.

Kuper R, Kröpelin S (2006) Climate-Controlled Holocene Occupation in the Sahara: Motor of Africa's Evolution. Science 313: 803-807.

Kusimba CM, Kusimba SB (2005) Mosaics and interactions: East Africa, 2000 b.p. to the present. In: Stahl AB (ed) African archaeology: A critical introduction. Blackwll publishing, Oxford, pp 394-419

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452. doi: 10.1093/bioinformatics/btp187

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt H-J, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308: 1034-1036. doi: 10.1126/science.1109792

Maxon RM (1994) East Africa An introductory history, second and revised edition edn. West Virginia University Press, Morgantown

Maxon RM (2009) East Africa: An introductory history. West Virginia University Press, Morgantown

McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433: 733-736.

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

Oliver R (1982) The Nilotic contribution to Bantu Africa. The Journal of African History 23: 433-442. doi: doi:10.1017/S0021853700021289

Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon J-M, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt H-J, Torroni A (2006) The mtDNA legacy of the levantine early upper palaeolithic in Africa. Science 314: 1767-1770. doi: 10.1126/science.1135566

Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi SQ, Thomas MG, Luiselli D, Bekele E, Bradman N, Balding DJ, Tyler-Smith C (2012) Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet 91: 83-96. doi: 10.1016/j.ajhg.2012.05.015

Pazzaglia A (1982) The Karimojong some aspects. EMI, Bologna

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.

Pereira R, Phillips C, Pinto N, Santos C, Santos SEBd, Amorim A, Carracedo Á, Gusmão L (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS ONE 7: e29684.

Phillipson DW (2005) African archaeology, 3rd edn. Cambridge University Press, Cambridge

Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo Á, Comas D (2004) Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Human Genetics 115: 439-447. doi: 10.1007/s00439-004-1164-0

Poloni ES, Naciri Y, Bucho R, Niba R, Kervaire B, Excoffier L, Langaney A, Sanchez-Mazas A (2009) Genetic evidence for complexity in ethnic differentiation and history in East Africa. Ann Hum Genet 73: 582-600. doi: 10.1111/j.1469-1809.2009.00541

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: 611-20. doi: 10.1016/j.ajhg.2010.02.014

Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, Sica L, Mouguiama-Daouda P, Comas D, Tzur S, Balanovsky O, Kidd KK, Kidd JR, van der Veen L, Hombert JM, Gessain A, Verdu P, Froment A, Bahuchet S, Heyer E, Dausset J, Salas A, Behar DM (2008) Maternal traces of deep common ancestry and asymetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. Proceedings of the National Academy of Sciences of the United States of America 105: 1596-1601.

Rhouda T, Martínez-Redondo D, Gómez-Durán A, Elmtili N, Idaomar M, Díez-Sánchez C, Montoya J, López-Pérez MJ, Ruiz-Pesini E (2009) Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait. Mitochondrion 9: 402-407.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt H-J (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67: 1251-1276.

Richards M, Macaulay V, Hill C, Carracedo A, Salas A (2004) The archaeogenetics of the dispersals of the Bantu-speaking peoples M. Jones, ed. Traces of Ancestry: Studies in honour of Colin Renfrew. McDonald Institute for Archaeological Research, Cambridge, pp 75-88

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A (2003) Extensive female-mediated gene flow from Sub-Saharan Africa into Near Eastern Arab populations. American Journal of human genetics 72: 1058-1064.

Rito T, Richards MB, Fernandes V, Alshamali F, Cerny V, Pereira L, Soares P (2013) The first modern human dispersals across Africa. PLoS ONE 8: e80031. doi: 10.1371/journal.pone.0080031

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. American Journal of human genetics 71: 1082-1111.

Scozzari R, Torroni A, Semino O, Cruciani F, Spedini G, Benerecetti SA (1994) Genetic studies in Cameroon: Mitochondrial DNA polymorphisms in Bamikeke. Human Biology 66: 1-12.

Silva M, Alshamali F, Silva P, Carrilho C, Mandlate F, Trovoada MJ, Černý V, Pereira L, Soares P (2015) 60,000 years interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. Scientific Reports 5:12526

Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Černý V, Pereira L (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. Molecular Biology and Evolution 29: 915-927. doi: 10.1093/molbev/msr245

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. American Journal of Human Genetics 84: 740-59. doi: 10.1016/j.ajhg.2009.05.001

Soares P, Rito T, Pereira L, Richards MB (2014) A genetic perspective on African prehistory. In: Jones SC, Stewart B (eds) Vertebrate Paleobiology and Paleoanthropology Book Series. Springer, (forthcoming)

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M, Pääbo S, Watson E, Risch N, Jenkins T, Kidd KK (1996) Global Patterns of Linkage Disequilibrium at the CD4 Locus and Modern Human Origins. Science 271: 1380-1387. doi: 10.1126/science.271.5254.1380

Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernandopulle N, Lema G, Nyambo TB, Ramakrishnan U, Reed FA, Mountain JL (2007) History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. Mol Biol Evol 24: 2180-2195.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM (2009) The genetic structure and history of Africans and African Americans. Science 324: 1035-44. doi: 10.1126/science.1172257 van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30: E386-94. doi: 10.1002/humu.20921

Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Paabo S (1996) mtDNA sequence diversity in Africa. American Journal of human genetics 59: 437-444.

Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. The American Journal of Human Genetics 61: 691-704.

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Clark Howell F (2003) Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 423: 742-747.

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF (2005) Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Europeab Journal of Human Genetics 13: 867-876.

Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13: 555-556.

Table 1. Shared haplotypes (SH) and admixture proportions between the four populations from Uganda and the six main African regions. Analysis was performed considering perfect matches between the source populations and the case-study population (P_0) or one- or two-mutational step differences between their haplotypes (P_1 and P_2, respectively)

		North Africa	West-Central Africa	Eastern Africa	South Africa	Southeast Africa	Southwest Africa
Bantu	SH	21	32	29	2	17	10
	P_0	0.1096 (0.0373)	0.4369 (0.0593)	0.3474 (0.0569)	0.0085 (0.0110)	0.0870 (0.0337)	0.0106 (0.0122)
	P_1	0.1354 (0.0349)	0.4585 (0.0509)	0.3282 (0.0479)	0.0046 (0.0069)	0.0561 (0.0235)	0.0171 (0.0132)
	P_2	0.1667 (0.0362)	0.4836 (0.0485)	0.2722 (0.0432)	0.0027 (0.0051)	0.0571 (0.0225)	0.0176 (0.0128)
Eastern Nilotic	SH	18	21	46	1	13	12
	P_0	0.0965 (0.0274)	0.2052 (0.0375)	0.6165 (0.0451)	0.0006 (0.0022)	0.0615 (0.0223)	0.0197 (0.0129)
	P_1	0.1082 (0.0226)	0.3161 (0.0338)	0.4764 (0.0363)	0.0019 (0.0032)	0.0701 (0.0186)	0.0273 (0.0118)
	P_2	0.1427 (0.0239)	0.3724 (0.0330)	0.3893 (0.0333)	0.0013 (0.0025)	0.0708 (0.0175)	0.0234 (0.0103)
Western Nilotic	SH	11	16	17	3	8	9
	P_0	0.0882 (0.0536)	0.4704 (0.0943)	0.3502 (0.0902)	0.0054 (0.0138)	0.0457 (0.0395)	0.0402 (0.0371)
	P_1	0.1009 (0.0470)	0.5647 (0.0774)	0.2740 (0.0697)	0.0041 (0.0100)	0.0327 (0.0278)	0.0236 (0.0237)
	P_2	0.1088 (0.0469)	0.5841 (0.0743)	0.2567 (0.0659)	0.0037 (0.0092)	0.0293 (0.0254)	0.0174 (0.0197)
Central Sudanic	SH	1	1	1	0	1	0
	P_0	0.2956 (0.4563)	0.6855 (0.4643)	0.0126 (0.1114)	0.0000 (0.0000)	0.0063 (0.0791)	0.0000 (0.0000)
	P_1	0.1758 (0.1439)	0.5988 (0.1853)	0.1502 (0.1351)	0.0004 (0.0077)	0.0735 (0.0986)	0.0012 (0.0133)
	P_2	0.1052 (0.1023)	0.5511 (0.1658)	0.2936 (0.1518)	0.0003 (0.0053)	0.0330 (0.0596)	0.0169 (0.1658)

Fig. 1 The mtDNA haplogroup distribution amongst 409 samples from Uganda

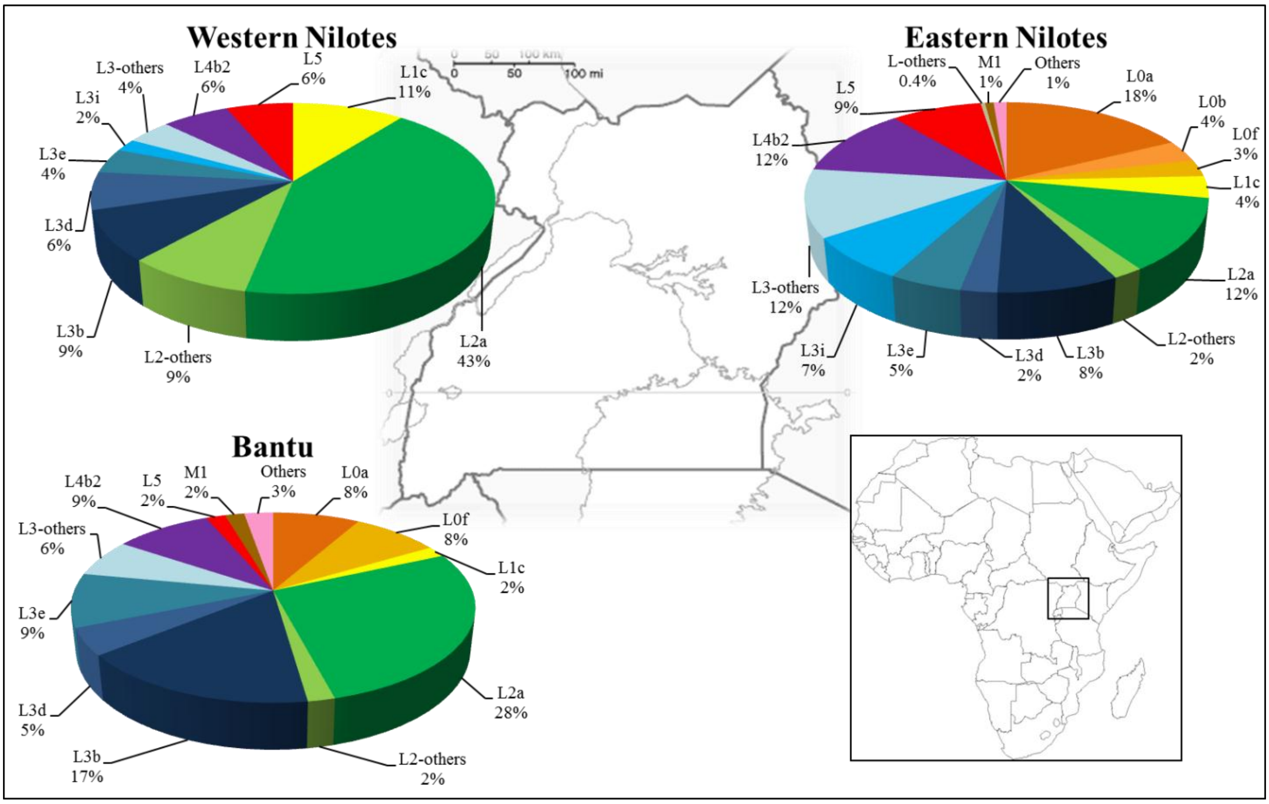


Fig. 2 PCA depicting the first two principal components of relationships between Ugandan linguistic groups and other populations of Africa, with Bantu distinguished from non-Bantu speakers

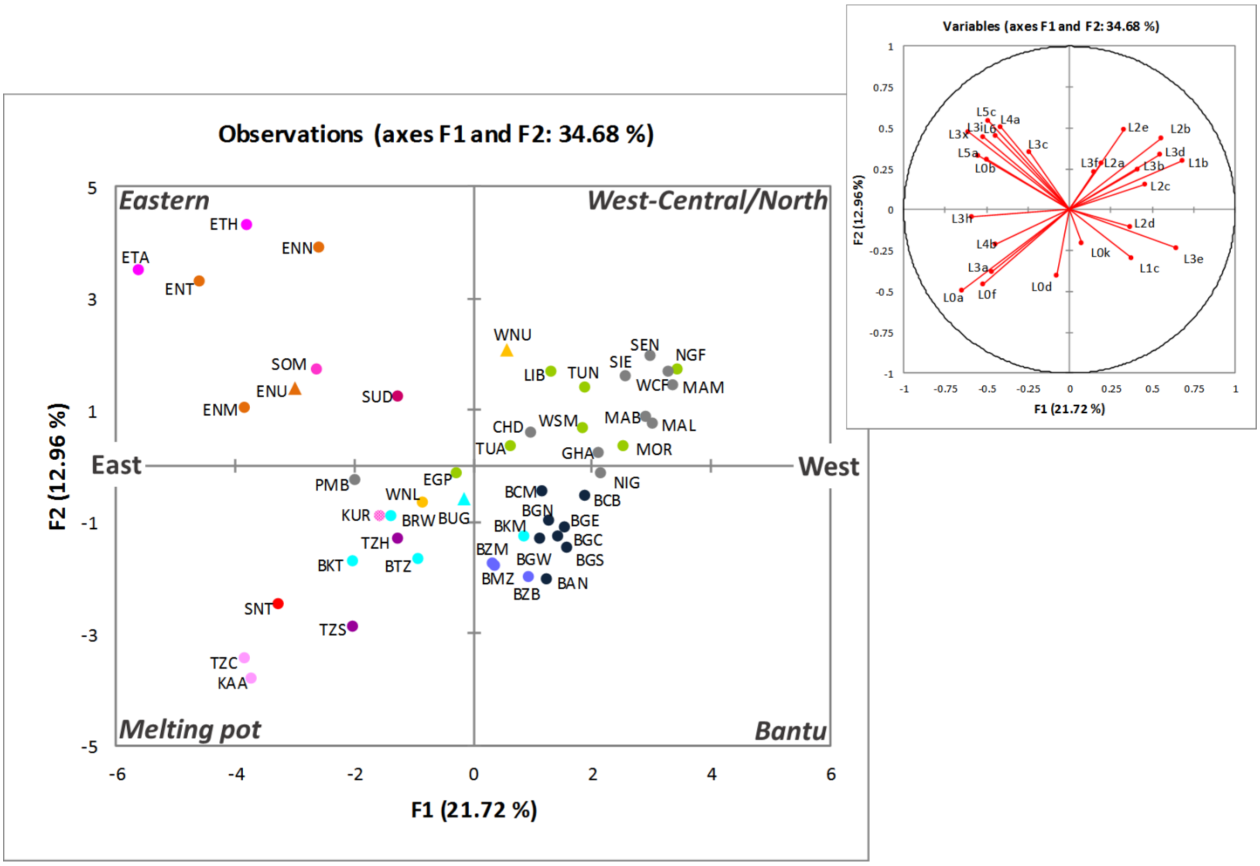


Fig. 3 Phylogenetic network of HVS-I variation in paragroup L3a'h'i'x. Links are labelled with the nucleotide position for each transition variant, less 16,000, with transversions indicated by the base change and reversions towards R root with suffix @. Major identifiable subclades are indicated, but note the weak phylogenetic resolution of HVS-I close to the root of L3 where different haplogroups cannot be distinguished

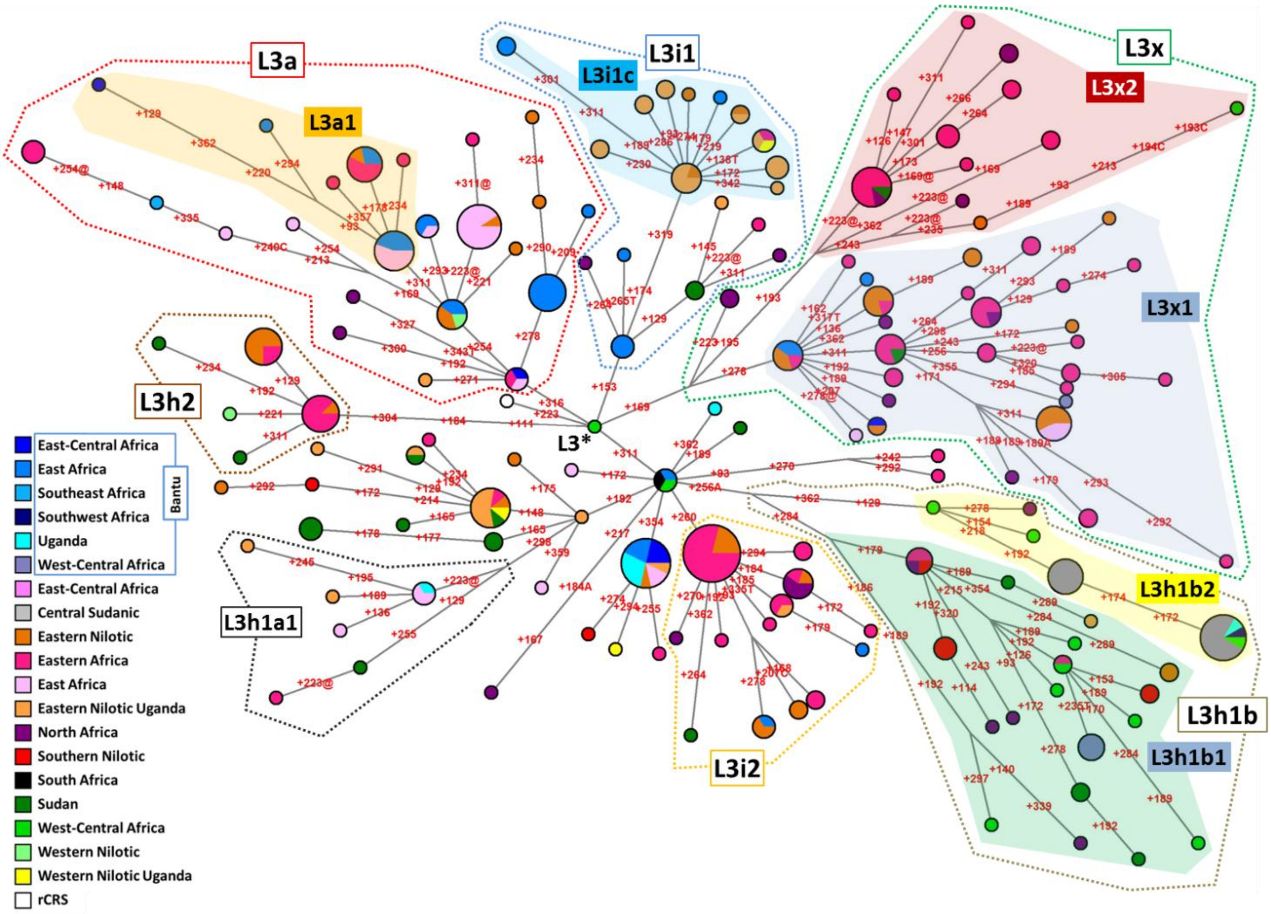
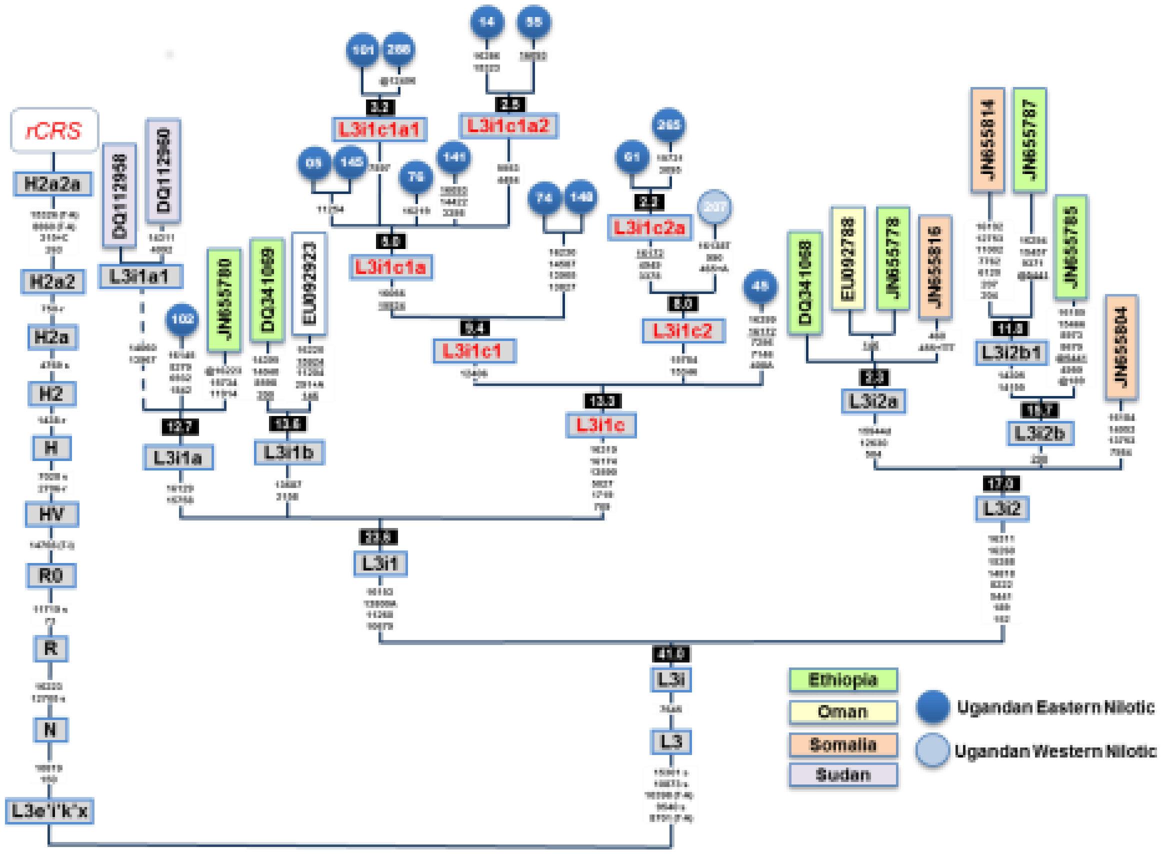


Fig. 4 Maximum-parsimony phylogenetic tree of whole-mtDNA genome variation in haplogroup L3i, with branches labelled with the nucleotide position for each transition variant, with transversions indicated by the base change, @ indicating a reversion, + indicating insertion of the following nucleotide. ML ages of the nodes in ka are indicated (see Table S4 for ρ estimates and 95% CIs). Dotted lines indicate lineages for which the control region was not tested. Tree rooted with the rCRS

Figure 4



Supplementary material

Supplemental files S1-S3, Tables S1–S4 and supplemental Figures S1–S11 are available online.