

The expansion of mtDNA haplogroup L3 within and out of Africa

Research Article

Pedro Soares¹, Farida Alshamali², Joana B. Pereira^{1,3}, Verónica Fernandes^{1,3}, Nuno M. Silva¹, Carla Afonso¹, Marta D. Costa^{1,3}, Eliska Musilova⁴, Vincent Macaulay⁵, Martin B. Richards³, Viktor Cerny⁶, Luísa Pereira^{*1,7}

¹IPATIMUP (Instituto de Patologia e Imunologia Molecular da Universidade do Porto), Porto 4200-465, Portugal

²Dubai Police GHQ - General Department of Forensic Sciences & Criminology, Dubai, United Arab Emirates

³Institute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

⁴Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic

⁵Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK

⁶Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague, The Czech Republic

⁷Faculdade de Medicina da Universidade do Porto, Porto 4200-319, Portugal

*To whom correspondence should be addressed:

Email: lpereira@ipatimup.pt

IPATIMUP, R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal

Phone: +351 22 5570700, Fax: +351 22 5570799

Running title: mtDNA haplogroup L3

Keywords: mtDNA; complete genomes; haplogroup L3; Out of Africa; modern human expansions

Abstract

Although fossil remains show that anatomically modern humans dispersed out of Africa into the Near East ~100–130 ka, genetic evidence from extant populations has suggested that non-Africans descend primarily from a single successful later migration. Within the human mtDNA tree, haplogroup L3 encompasses not only many sub-Saharan Africans but also all ancient non-African lineages, and its age therefore provides an upper bound for the dispersal out of Africa. An analysis of 369 complete African L3 sequences places this maximum at ~70 ka, virtually ruling out a successful exit before 74 ka, the date of the Toba volcanic super-eruption in Sumatra. The similarity of the age of L3 to its two non-African daughter haplogroups, M and N, suggests that the same process was likely responsible for both the L3 expansion in Eastern Africa and the dispersal of a small group of modern humans out of Africa to settle the rest of the world. The timing of the expansion of L3 suggests a link to improved climatic conditions after ~70 ka in Eastern and Central Africa, rather than to symbolically mediated behavior, which evidently arose considerably earlier. The L3 mtDNA pool within Africa suggests a migration from Eastern Africa to Central Africa ~60–35 ka, and major migrations in the immediate postglacial, again linked to climate. The largest population size increase seen in the L3 data is 3–4 ka in Central Africa, corresponding to Bantu expansions, leading diverse L3 lineages to spread into Eastern and Southern Africa in the last 3–2 ka.

Introduction

Since the emergence of *Homo sapiens* 150–200 thousand years (ka) ago, the species is thought to have lived in small, scattered populations and been at risk of extinction for much of its existence (Newman 1995; Willoughby 2007). This low effective population size is reflected in deeper parts of the mitochondrial DNA (mtDNA) tree of modern humans. Although the tree is highly starlike at shallower time depths, suggesting numerous episodes of rapid growth in the human population in the more recent past, it is only at a third of the time depth of the entire tree with the emergence of the L3 haplogroup that the first multifurcating node occurs (seven sub-branches, including haplogroups M and N that encompass between them all the ancient diversity observed outside Africa) (Behar et al. 2008; Torroni et al. 2006; Watson et al. 1997). The explanation for this dramatic early population expansion after the appearance of L3 is unknown, but advances in technological or cognitive capacities and response to climatic change have been suggested (Forster 2004; Klein 1992; Mellars 2006; Scholz et al. 2007). Two main routes out of Africa are still widely discussed: the Sinai, linking Egypt to the Levant, and the Bab el-Mandab strait, from the Horn of Africa to the Arabian Peninsula (Derricourt 2005). Early dates for the settlement of Southeast Asia by ~50 ka (Barker et al. 2007) and Australia by 48 ka (Turney et al. 2001) combined with the distribution and ages of mtDNA lineages (Atkinson et al. 2008; Kivisild et al. 2003; Macaulay et al. 2005; Soares et al. 2009; Thangaraj et al. 2005) have suggested a “southern coastal route” via Arabia (Field and Lahr 2005) as the sole major exit from Africa in the Late Pleistocene (Beyin 2006; Bulbeck 2007; Mellars 2006; Richards et al. 2006). Nevertheless, controversy about the timing in particular remains.

It is usually assumed that the exit should take place during an interglacial period, implying either marine isotope stage 5 (MIS 5 at 130–75 ka) or MIS 3 (60–25 ka) (Bar-Yosef 1995). An early arrival in South Asia (as early as 125–130 ka) has recently been proposed on the basis of lithic evidence in India and the United Arab Emirates (Haslam et al. 2010), with tool kit affinities between Arabia and the late Middle Stone Age (MSA) of northeast Africa suggesting (although not to all: Lawler 2011) that low sea level and increased rainfall during the transition between MIS 6 and 5, rather than technological innovation, might have allowed this proposed early dispersal into Arabia (Armitage et al. 2011). This would imply that the early movement of modern humans into the Levant at 100–130 ka known from fossil evidence (Grün et al. 2005) led to further expansions into Asia. Several other models propose an exit from Africa prior to the Toba super-eruption in Sumatra at 73.5 ka. For example,

Oppenheimer (2003) has proposed an exit towards the end of stage 5, due to a fall in the salinity of the Red Sea. On this scenario, early modern Asians might have been profoundly affected by the Toba eruption; although continuity of toolkits across the Toba layer in India has also been proposed (Petraglia et al. 2007).

Cohen et al. (2007) have argued for a similar model, providing evidence for several “mega-droughts” across tropical Africa 130–90 ka, and proposing that the dispersal of modern humans took place during stage 5, 90–70 ka, albeit along the Nile valley, in the crossover period before high aridity hit Eurasia after 70 ka. However, Scholz et al. (2007) argued from the same data that the stable moist phase in tropical Africa during the glacial period of MIS 4, beginning ~70 ka, most likely coincided with the expansion of modern human populations, fitting a spread from Africa a little later, by ~60 ka. Mellars (2006) also invoked environmental change and a similar expansion time, but coupled this to innovations related to behavioural modernity.

The time window for the out-of-Africa migration on the basis of mtDNA lies between the emergence of haplogroup L3 in Eastern Africa and its derivative non-African haplogroups M and N, which most likely arose during the departure or outside Africa (Richards et al. 2006). Since the age of L3 places an upper bound on the time of the out-of-Africa migration, a sufficiently precise estimate would allow us to test these hypotheses – at least to the extent of determining whether maternal lineages from any putative pre-Toba dispersal might have survived to the present day. The mtDNA molecular clock estimates have hitherto, however, indicated a range for the time of the most recent common ancestor (TMRCA) of haplogroup L3 at around 50–90 ka (Behar et al. 2008; Endicott et al. 2008; Macaulay et al. 2005; Salas et al. 2002; Soares et al. 2009), placing a rather imprecise upper bound on the timing of the dispersal. An even younger date for the exit better fits estimates from Y-chromosome variation (Shi et al. 2010), but there is considerable uncertainty about Y-chromosome mutation rates (Zhivotovsky et al 2004).

The population expansions after the emergence of haplogroup L3 most likely led not only to global colonisation but also to range expansions within Africa. There is an increase in microlithic/Mode 5 (Later Stone Age or LSA) technologies in the archaeological record from ~50 ka onwards throughout Africa, but this is uneven and gradual at best for much of the continent (Barham and Mitchell 2008; Phillipson 2005). Climatology has suggested more recent episodes of climate change in Africa that might have led to population movements and expansion. For example, following the Last Glacial Maximum (LGM) ~21 ka and the Younger Dryas cold snap, moist conditions prevailed in West Africa from 11.5–5.1 ka

(Weldeab et al. 2007), and in the Sahara the Holocene Climatic Optimum lasted until 7.3 ka (Kuper and Kröpelin 2006), with a shift to the present aridity from ~6 ka (Brooks et al. 2005). These early Holocene conditions likely influenced the spread of pastoralism into the Sahara (Brooks et al. 2005; Kuper and Kröpelin 2006). One major expansion in Africa that left a reasonably clear archaeological, linguistic and genetic trail is the spread of Bantu speakers, alongside early domesticated resources and subsequently iron-working, that began in West-Central Africa by ~3.5 ka, resulting in continent-wide dispersals both south and east (Phillipson 2005).

Eastern Africa, thought to be the most genetically diverse region of the world and the likely place of origin of L3 (Torroni et al. 2006), as well as the source for non-African populations (Tishkoff et al. 2009), is clearly pivotal to most of these expansions (Scheinfeldt et al. 2010; Campbell and Tishkoff 2011). There is also evidence that at least part of this region acted as a refuge during the severest climatic episodes of the last several hundred thousand years, particularly around wooded lake margins and perhaps also in coastal regions (Barham and Mitchell 2008; Basell 2008) – although Compton (2011) has recently suggested that the Mediterranean coast and the South African coastal plain may also have acted as refugia for humans during this period. Moreover, Eastern Africa was a key secondary dispersal area for the Bantu expansion (Phillipson 2005; Salas et al. 2002). We have therefore enriched the databases of mtDNA L3 diversity in Eastern Africa, by firstly characterising at low resolution 327 samples from Sudan, Ethiopia and Somalia, and then selecting from these 70 L3 lineages for complete mtDNA genome sequencing. The application of several methods and mutation rates for the evaluation of TMRCAs allowed us to narrow the time span for the emergence of L3 and establish an upper bound for the out-of-Africa migration. In addition, Bayesian skyline analysis of 328 complete L3 sequences and founder analysis of 2,359 L3 hypervariable segment I (HVS-I) sequences enabled us to infer both local demographic expansions and migrations within Africa.

Materials and Methods

Samples, mtDNA sequencing and haplogroup affiliation

We collected a total of 102 Sudanese, 77 Ethiopian (both emigrants in Dubai) and 148 Somali (refugees in Yemen) samples, belonging to unrelated individuals, who gave appropriate informed consent for their biological samples to be used for mtDNA characterization. The work complied with the Helsinki Declaration of Ethical Principles (59th WMA General Assembly, Seoul, October 2008) and was approved by the Ethics Committee of the University of Porto (11/CEUP/2011). We sequenced hypervariable segments I and II (HVS-I and HVS-II) of all Sudanese and Ethiopian samples, and HVS-I in the Somali population, using a procedure described previously (Pereira et al. 2000). This information was used to assign samples to haplogroups, following the most recent phylogenetic evidence, reported in Phylotree website (van Oven and Kayser 2009). The sequences obtained are reported in Supplemental Table S1 and Supplemental Material 2.

We selected for complete mtDNA sequencing a total of 21 Sudanese, 16 Ethiopian and 20 Somali samples chosen from the sequences characterised in this work, and 11 from Chad and two from Soqotra belonging to haplogroup L3 from datasets published previously (Cerný et al. 2009a; Cerný et al. 2007). We followed the methodology and checking procedures reported in Pereira et al. (2007), and mutations were scored relatively to the revised Cambridge reference sequence (Andrews et al. 1999). The 70 complete mtDNA sequences are deposited in GenBank (accession numbers JN655773-JN655842).

We collected a database of published HVS-I and complete L3 lineages that are described in Supplemental Text, Table S2 and S3.

Statistical analyses

For the L3 phylogeny reconstruction, preliminary reduced-median network analyses (Bandelt et al. 1995) led to a suggested branching order for the trees, which we then constructed most parsimoniously by hand. We used the software mtDNA-GeneSyn (Pereira et al. 2009) to convert files.

For estimation of the TMRCA for specific clades in the phylogeny, we used the ρ statistic (Forster et al. 1996) and maximum likelihood (ML). We used ρ (the mean sequence divergence from the inferred ancestral haplotype of the clade in question) with a mutation rate estimate for the complete mtDNA sequence of one substitution in every 3,624 years, correcting for purifying selection using the calculator provided (Soares et al. 2009) and a

synonymous mutation rate of one substitution in every 7884 years. We estimated standard errors as in Saillard et al. (2000). We also obtained ML estimates of branch lengths using PAML 3.13 (Yang 1997), assuming the HKY85 mutation model with gamma-distributed rates (approximated by a discrete distribution with 32 categories). We converted mutational distance in ML to time using the same complete mtDNA genome clock.

We obtained Bayesian skyline plots (BSPs) (Drummond et al. 2005) from BEAST 1.4.6 (Drummond and Rambaut 2007) for the complete L3 sequences with a relaxed molecular clock (Drummond et al. 2006) (lognormal in distribution across branches and uncorrelated between them) and the HKY model of nucleotide substitutions with gamma-distributed rates. Previous maximum likelihood analyses indicated that HKY is an adequate model and it is not necessary to use more complex models (Soares et al 2009). BSPs estimate the effective population size through time using random sequences from a given population.

Haplogroup L3 clearly does not equate to population data, but the signal associated with this haplogroup might nevertheless signal demographic processes in the populations carrying it, as suggested previously (Atkinson et al. 2009). The main purpose of this analysis was to provide BSP plots based on the L3 data – not to perform an independent calibration of the L3 phylogeny, since we did not discern any secure calibration point within L3. With this purpose, we performed an analysis aiming to approximate the mutation rate to the one used in the remaining analyses. This was based on the age of L3, assigning a normal distribution with a mode age of 65 ka and a 95th percentile varying between 60–70 ka, based on the age estimations obtained with ρ and ML. Since the mutation rate is time-dependent we provided minimum age estimates for some of the younger clades, maintaining some level of independence from the remaining analyses with ρ and ML. Some of the subclades within L3d1a1, L3e1a, L3e1b, L3e1d, L3e3a, L3e4a, L3f1b1 and L3f1b4 that are present in Southern Africa (and Central Africa) were assumed to have arrived there due to Bantu migrations (Newman 1995), a notion bolstered by the founder analysis results (see below). In this way, the age attributed to these clades was at least 3.5 ka old, since the clade must be at least the age of the Bantu expansion, but could be older if it already carried some diversity in Central Africa before the beginning of the expansion southwards. To take this into account we set an exponential distribution with a mode of 3.5 ka and 95th percentile 8 ka, allowing the age to be substantially higher than 3.5 ka. BEAST uses a Markov-chain Monte-Carlo (MCMC) approach to sample from the posterior distributions of model parameters (branching times in the tree and substitution rates). Specifically, we ran 100,000,000 iterations, with samples drawn every 10,000 MCMC steps, after a discarded burn-in of 10,000,000 steps. We checked

for convergence to the stationary distribution and sufficient sampling by inspection of posterior samples and we visualized BSPs with Tracer v1.3. We used a generation time of 25 years, as in Fagundes et al. (2008), and forced the larger subhaplogroups to be monophyletic in the analysis since we aimed at a tree structure that is directly comparable with the remaining analyses.

We used the mutation rate obtained in the L3 overall tree to run analyses for the subhaplogroups and sub-regions. The subhaplogroup analyses allowed to us distinguish which clades were mainly responsible for the increases in effective population sizes observed in the overall L3 analysis. The sub-regions analyses permit to observe if a given region contained the signal for a given population increase in its specific L3 sequences, indicating a possible population expansion within that region involving several of the subclades. In order to perform a systematic comparison and description of the increment periods in the effective population size of the BSP, we calculated a rate of population size change through time and it was calculated relatively to individuals per 100 individuals for a period of 100 years (see figure S1 for a graphic of the variation of this value for the BSP of overall L3). As a definition we considered the rate of increase of at least 1 individual per 100 individuals in 100 years as a steeper increase, matching well the visualization of the BSPs.

To visualize the geographical distribution of L3 and its subhaplogroups, we constructed interpolation maps using the “Spatial Analyst Extension” of ArcView version 3.2 (www.esri.com/software/arcview/). We used the “Inverse Distance Weighted” (IDW) option with a power of two for the interpolation of the surface. IDW assumes that each input point has a local influence that decreases with distance. The geographic location used is the centre of the distribution area from which the individual samples of each population were collected. The data used are listed in Supplemental Table S2.

We employed a founder analysis (Richards et al. 2000) for the L3 sequences in Africa. It assumes a strict division between source and sink populations and a series of criteria to partly account for homoplasy and back migrations between regions. The distinction between source and sink populations is straightforward for most lineages in Africa, except for some deep-ancestry clades which are present in both Eastern Africa and Central Africa. Some cautionary steps were taken in order to allow for the effect of gene flow in both directions. We performed the following founder analyses:

(1) *From Eastern Africa into Central Africa:* Eastern Africa was the source of most of the ancient L3 variation, although some subclades (L3b, L3d and L3e, as we will discuss in Results) most probably emerged in Central Africa, with their presence in Eastern Africa

related to migrants from Central Africa, and not the other way around; so these subclades were not assumed to be founders from Eastern Africa. Instead two founders (containing the entire L3bd and L3e clades) were accounted for as Eastern African founders into Central Africa. We opted for enlarged definitions of Central Africa (including all of Nigeria) and East Africa (extending from Sudan to Tanzania) for the four founder analyses, to minimize potential problems with genetic drift that might arise with selected smaller datasets. In this case the inclusion of all of Nigeria in Central Africa would only increase the dataset of the sink population that is perfectly valid when considering an eastern source.

(2) *From Central Africa into Eastern Africa:* Central African haplogroups were involved in important expansions in Central Africa, namely the Bantu expansion. In this case, only L3b, L3c and L3e would provide founders into Eastern Africa. We included a single founder, the L3 root type that emerged and evolved *in situ*, to account for the remaining L3 variation observed in Eastern Africa. Although only a small fraction of Nigeria was a source for Bantu-speakers, the extended boundaries are nevertheless beneficial for its use as a source for Bantu lineages into Eastern Africa (or Southern Africa) since it would either contain lineages that had subsequently moved from the more strictly defined Central Africa, improving the source dataset, or contain lineages not involved in these migration that would not impact on the analysis.

(3) *From Central and Eastern Africa into North Africa:* We excluded the Sahel from the analysis as this region has had continuous recurrent gene flow both from north and south, dissolving the definition of well-defined populations and the populations in this area would be difficult to either classify as source or sink population.

(4) *From Central and Eastern Africa into Southern Africa:* We performed this analysis in two ways, considering two versions of the Southern African dataset as the sink population. At least two major routes are thought to have been taken by Bantu speakers spreading into Southern Africa, a “western stream” and an “eastern stream”. The western stream took the expanding Bantu from eastern Nigeria/Cameroon into Angola, South Africa and Botswana from around 3.5 ka ago, whereas the eastern stream reached Mozambique and South Africa within the last 2 ka, starting from a core area in the Great Lakes region of Uganda 2.5 ka ago (Bandelt et al. 2001; Phillipson 2005; Salas et al. 2002). In the first, we considered the entire Southern African dataset, but for the second we tested for the possibility of a more recent ancestry for populations at the extreme southeastern of the continent, which likely received migrants primarily from the “eastern stream” of the Bantu expansion as well as more recent

immigration from Eastern Africa (Huffman 2007; Phillipson 2005). Here we therefore only included samples from Mozambique and South Africa.

We reconstructed, haplogroup by haplogroup, HVS-I networks in the range 16051–16400 bp of the reference sequence (Andrews et al. 1999). Data used are indicated in Table S2. We used an *fI* criterion (which dictates that a sequence type is only considered a founder if it presents derived variation in the hypothetical source population) to identify founders, in order to allow for some homoplasy and back migrations (Richards et al. 2000).

We estimated the age of the migration of each founder using the ρ statistic (Forster et al. 1996) and an HVS-I mutation rate of one mutation every 16,677 years (Soares et al. 2009). In order to assess the error in the Bayesian partitioning across the different migration times realistically, we calculated an effective number of samples in each founder. This was obtained by multiplying the number of samples in each founder by a ratio of the variance assuming a star-like network and the variance calculated with the Saillard et al. method (Saillard et al. 2000). We scanned the distribution of founder ages for each region defining equally spaced 200-year intervals for each migration from 0 to 100 ka. We performed a second analysis using defined migration times, based on the previous scan in the context of archaeological and climatological evidence. This second analysis allowed us to fractionate the L3 lineages into each migration and to probabilistically detect which lineages moved in each migration.

Results

Phylogeography of haplogroup L3

The complete mtDNA phylogeny of haplogroup L3 is shown in the Supplemental Tree, including age estimates using the complete genome (using ρ and ML) and the synonymous (using ρ) clocks. Figure 1 presents an outline topology of L3, indicating the primary branches with age estimates using the same methods, scaled against the ML estimates.

L3 most likely had an origin in Eastern Africa (Torroni et al. 2006). This is supported by the presence of all major branches, with L3a and L3h virtually specific to the region and L3eikx and the L3f haplogroups having a probable origin there as well, while for L3bcd the region of origin is less clear and will be discussed in more detail below. The other two main branches of L3, M and N, exist only outside Africa, except for some back-migrations into Africa around 50–30 ka ago in the form of haplogroups U6 and M1 (Olivieri et al. 2006; Pereira et al. 2010) and some more recent intrusions (Cherni et al. 2009; Ottoni et al. 2010). Since there is strong evidence that the dispersal out of Africa was via the Horn, soon after L3 arose (Macaulay et al. 2005), the distribution of M and N also point to Eastern Africa as the centre of gravity for L3. It seems likely that L3 dates somewhere between 60–70 ka, as the TMRCA estimates vary between 58.9 ka (using ρ and the complete genome) and 70.2 (using ML), with the synonymous clock providing a value between the two (63.1 ka). We checked if the L3 tree rejected a strict molecular clock by running the ML analysis without stipulating a molecular clock and performing a likelihood ratio test that clearly did not reject the clock hypothesis ($p=0.9899$). These results led us to consider an age of 65 ka (varying between 60 and 70 ka in the 95% percentile) in the internal calibration of BEAST, as there was no other reliable calibration point we could use. The ages using the complete genome and the synonymous clock were 1.04 and 0.97 times the BEAST estimates respectively, indicating that the BSPs were calculated using a similar rate to the other analyses, as we intended.

The BSP for L3 overall points to three main episodes of population growth within Africa associated with L3 (Figure 2 and Table 1). The first is steepest at ~40 ka although with a gradual start from before 50 ka. This signal is not observed in the BSPs of any of the individual L3 branches analysed; it corresponds to the emergence of the various subhaplogroups between 40–50 ka. Although the root type of L3 probably expanded soon after its emergence 60–70 ka, given its seven basal branches, the single coalescence point was

not detected in the BSP. The second episode is a steep increase in the effective population size around the beginning of the Holocene (Table 1), associated particularly with the main Central African haplogroups (L3bd, L3e). It is also suggested by several other haplogroups present in Eastern Africa at the time (L3h, L3x, L3f); no signal was detected for each of them individually (Figures S2–S5), but it was observed when they were combined (Figure S6). The third major expansion took place from ~4 ka and led to the largest increase (Table 1) and as the signal is only pronounced in the Central African BSP (Figure S8), this seems to have been a localized occurrence. Atkinson et al. (2009) performed a similar analysis on African lineages, including haplogroup L3, but the resolution of L3 was much lower in that case (80 sequences against >300 now) and they did not detect the two distinct increases in the Holocene.

Eastern African origins

The diversification of L3 in Eastern Africa began early, as demonstrated by the ages of L3a and L3h (Figure 1), both of which are virtually specific to this region (Figure 3A). The BSP for Eastern Africa (Figure S6) alone rises most steeply only after 40 ka (Table 1), but the plot shows a progressive increase from before 50 ka. Accordingly, the scan of HVS-I diversity of founder L3 lineages in Eastern Africa showed a peak at ~58.8 ka (corresponding to nearly three-quarters of the L3 data in Eastern Africa; Table 2), followed by a second peak at ~1.8 ka.

Haplogroup L3f most likely also arose in Eastern Africa, where it is both most frequent (Figure 3B) and most diverse (Salas et al. 2002), although several branches were carried by migrants into the Sahel and Central Africa: L3f1 has two main subclades (L3f1a and L3f1b) with basal lineages from Eastern Africa; L3f2 is specific to Eastern Africa; L3f3 seems to have expanded into the Sahel 8–9 ka ago hypothetically from an Eastern African source (Cerný et al 2009b), as it is also present in Central and North Africa. The scan of the founder lineages in Central Africa (Figure 4B) suggests that this region received a strong genetic input of L3f lineages in the beginning of the Holocene, around the time of the expansion of L3f3, perhaps reflecting a wide scale demographic process in the early Holocene, also suggested from the BSP analysis (Figure S6, Table 1). L3eikx also most likely originated in Eastern Africa, with subclades L3i and L3x virtually exclusive to this region (Figure 3A) and the rare L3k mainly present in North Africa. The complete genome tree confirms that L3e (the most frequent African L3 clade: Figure 3C), however, seems to have an origin in Central/West Africa (Bandelt et al. 2001; Salas et al. 2002): the Eastern

African samples are clearly derived within the tree, suggesting gene flow from Central Africa. The recent gene flow into Eastern Africa is represented by the peak at ~1.8 ka (Figure 4A, Table 2) in the HVS-I founder analysis, which includes some L3e lineages.

L3bcd has three main subclades, with L3b and L3d tentatively united by a transition at control-region position 16124 to form the putative subclade L3bd. The great age of L3bcd and its wide distribution across Africa makes phylogeographic inferences difficult. Furthermore, L3c is extremely rare: only two samples have been detected so far, one in Eastern Africa and the other in the Near East. This might echo an early origin of L3bcd in Eastern Africa, before moving west, but its rarity makes this conclusion extremely tentative. In a scenario of an early origin of L3bcd in Eastern Africa, M and N would be the only subclades of L3 to have most likely originated outside of Eastern Africa (although an origin in Eastern Africa remains possible: Richards et al. 2006). L3b and L3d most likely began to diversify in Central/West Africa, representing the earliest major spread of L3 lineages within Africa that we are able to detect. The recent peak of the founder scan for L3, dating to ~1.8 ka in Eastern Africa, mainly comprises L3b and L3d lineages in the corresponding partition in Table 2 (founders F8, F16, F17, F25 in supplemental Table S4). Many of these lineages are also present in Southern Africa, with founder age estimates indicating they arrived very recently, probably again within the last two millennia. The frequency of these four lineages is higher in the southern part of Eastern Africa (9% in Tanzania, 4.7% in Somalia, 3.2% in Kenya) than to the north (2.3% in Ethiopia, 0.78% in Sudan). Together with the age estimates, this points to a genetic input from West/Central Africa into Eastern Africa within the last few thousand years, into regions which now have many Bantu-speaking populations.

Central and West Africa

The most common L3 lineages in Central and West Africa are L3b, L3d, L3e and L3f and, apart from the latter, they most probably all originated within this region. L3b has a rare sub-branch (L3b2) and one wide-spread sub-branch (L3b1) with the frequency focus in Central/West Africa (Figure 3D). L3b1 has point estimates between 12.6–17.6 ka (depending on the clock) and its most common subclade, L3b1a, has point estimates of 11.7–14.8 ka, with star-like patterns suggesting involvement in major expansion. For L3d, all Eastern African samples except one (sample 78 in the Supplemental Tree) are derived in the tree. The HVS-I data also suggest that the diversity in Eastern Africa (23.1 ± 10.5 ka) is lower than the overall value for L3d (28.9 ± 7.4 ka) and for Central/West Africa (29.7 ± 7.1 ka), and again the focus of the frequency distribution (Figure 3E) is Central/West Africa.

Although L3eikx arose in Eastern Africa, L3e itself most probably had an origin in Central Africa where all five of its subclades are found. The ancestors of haplogroup L3e may have moved into Central Africa around the same time as for L3bd, as part of the same range expansion, especially if we consider an origin of L3bcd in Eastern Africa. In the founder analysis scan, L3bd and L3e were displayed as a single peak (although not very sharp) dating to ~42.4 ka. This is also the time where growth is observed in the BSP of the L3 data (Figure 2 and Table 1). Using this value as a migration time in the founder analysis, it would comprise three-quarters of the L3 Central African data. The second peak dating to 9.2 ka (Table 2) corresponds mainly to L3f1 lineages (founders F59 and F69, in Table S4), closely matching the time of expansion of L3f3 into the Sahel (Cerný et al. 2009b).

The BSPs for Central Africa show a steep rise in the effective population size after 4 ka and centred at 2.6 ka (Table 2). The signal detected in the BSP corresponds mainly to population growth within the Central African populations, not detected in the founder analysis, which only detects range expansions from one region to another. This demographic expansion most probably corresponds to the Bantu agriculturalist expansion that occurred after ~4 ka in Central Africa.

North Africa

The scan of founder lineage variation (Figure 4C) in North Africa indicates two peaks at 0.6 ka and at 6.6 ka. The dataset for North Africa includes a high fraction of direct matches (nearly 20%) which contribute to the more recent peak, most likely reflecting movements into North Africa from Central/Eastern Africa within the last 1000 years, perhaps including lineages carried with the trans-Saharan slave trade (Harich et al. 2010). We performed a founder analysis stipulating three migration times, including a third one of 35.0 ka, based on the ages of U6, L3k and the population increase in the BSP (Figure S7) with results presented in Table 1. The biggest slice corresponds to the peak at 6.6 ka, corresponding to one-third of the L3 lineages (Table 2), mainly affiliated to haplogroup L3e5, which is largely restricted to Northwest Africa. It dates to 12.4–13.6 ka, and so may have begun earlier than 6.6 ka. The other major lineages contributing to the 6.6 ka partition (Central African L3b, L3e1 and L3e2 lineages: founders F17, F28, F41 in Table S4), suggest the post-glacial period was characterized by gene flow across the Sahel belt (Cerný et al. 2007); these founder clades are mainly restricted to Northwest Africa, and absent from Egypt.

The BSP for North Africa also indicates the major rate of increase in the effective population size centered at about 10 ka (Figure S7, Table 2), contemporaneous with Central

and Eastern Africa. Since the majority of L3 lineages in the scan had an origin elsewhere, the timing of the growth signal observed in the North African BSP might in fact be describing the demography of the source, with many of the coalescences occurring prior to the dispersal into North Africa. However, an independent analysis of North African U6 (Pereira et al. 2010) indicated growth associated with this haplogroup ~10 ka.

Southern Africa

Southern African samples, all from Bantu speakers, are scattered throughout the L3 tree, with typically only a single sequence per founder type. A BSP for the Southern Africa data would be quite misleading, since these sequences consistently coalesce outside Southern Africa, and are few in number (19). Therefore, we must turn primarily to the larger volumes of HVS-I data.

Southern Africa is known to have received a high genetic input from further north with the Bantu expansion of the last two millennia (Pereira et al. 2001; Salas et al. 2002). The founder analysis for Southern Africa as a whole yielded a single peak at 3.2 ka (Figure 4D), closely matching the time of the initial expansion south of Bantu-speakers ~3.5 ka (Barham and Mitchell 2008). This indicates that all the L3 lineages in Southern Africa are likely to be the result of migrations within the last few millennia: L3 in the south does not predate the Bantu dispersals. Although the western stream was at least partly distinct from the eastern stream, the founder types are mostly the same, indicating that they were probably from the same proximal source in Central Africa or were already in contact before reaching the south. The founder analysis using only the data from Mozambique and South Africa yielded a single, younger peak of founder lineages at 1.8 ka (Figure 4D), matching closely the time of arrival of the “eastern stream” of Bantu-speaking agriculturists at the tip of the continent (Phillipson 2005), which probably predominated in this region (Huffman 2007). This peak also provides a good corroboration for the HVS-I clock employed here, also recently supported by estimates of the expansion into Remote Oceania (Soares et al. 2011).

Discussion

The time of emergence and expansion of L3 is a crucial issue in the study of human evolution since it provides an upper bound on the time of the out-of-Africa migration. Haplogroups M and N, the non-African branches of L3, most likely arose outside Africa or *en route*, serving as a lower bound for the dispersal. The range we have obtained for the age of L3 (~60–70 ka, with overlapping 95% confidence intervals for the various estimates from 59–79 ka) serves to virtually exclude the suggestion of an out-of-Africa exit (at least, of the ancestors of the non-African human populations alive today) before 73.5 ka, the time of the Toba super-eruption. Some archaeological data suggest the possibility of earlier migrations out-of-Africa during MIS5 or in the transition to it both through the Levant (Bar-Yosef 1992) or via the southern route (Armitage et al. 2011; Petraglia et al. 2007), but no maternal descendents of these possible earlier migrants remain in the extant population; L3 and its M and N daughters had not yet been born at that time.

Furthermore, since the age of the M and N Eurasian founders of ~50–65 ka (Soares et al. 2009) is close to the age of their ancestral L3 clade in Africa, the out-of-Africa dispersal was probably of a piece with the initial diversification and expansion of L3, so that the L3 expansions in Eastern Africa and the exit of modern humans from Africa ~60 ka were all part of a single demographic process. It seems likely that the moister climate after ~70 ka in Eastern Africa allowed dramatic human population growth (Scholz et al. 2007), perhaps associated with improved hunting, marine exploitation, exchange networks and possibly even plant food management strategies as suggested by Mellars (2006). This generated the oldest major signal of expansion in the human mtDNA tree, the radiation of L3, leading rapidly to the spread of *Homo sapiens* towards the rest of the world. It is worth stressing that this signal is not reflected in other mtDNA lineages at this time (Behar et al. 2008). Within Africa, the Pleistocene migrations detected in the L3 pool were responsible for the introduction of L3bd and L3e into Central Africa in the period between 60–35 ka (Figure 5A), but none reached Southern Africa at that time.

The timing therefore suggests that the expansion and dispersal may have been primarily stimulated by environmental factors (Cohen et al. 2007; Scholz et al. 2007), with the model of Scholz et al. (2007) providing the best current fit to the mtDNA chronology. This resembles the more detailed model proposed by Mellars (2006), but with the caveat that whilst Mellars proposed environmental change as an important driver, he also coupled the expansion to a step-change in the move towards behavioral modernity, by analogy with the

European Upper Paleolithic, and in contrast to more incremental models of the African transition to modernity. This increased evidence for modernity is most evident in the MSA finds in Southern Africa ~70–80 ka (*e.g.* Henshilwood et al. 2002, 2009; Texier et al. 2010), although it might plausibly have originated in Eastern Africa a little earlier (Mellars 2006).

However, neither an origin of L3 in Southern Africa, as suggested by Compton (2011), nor a dispersal from Eastern Africa to Southern Africa at ~70–80 ka, as implied by Mellars (2006), can be plausibly coupled with the expansion of haplogroup L3 as reconstructed here. The nearest phylogenetic relatives of L3, L4 and L6, are both localized entirely within Eastern Africa, Arabia and the Near East (Kivisild et al. 2004; Torroni et al. 2006; Behar et al. 2008), strongly supporting an origin for L3 in Eastern Africa, and effectively ruling out the possibility of a dispersal from Southern Africa, where the indigenous, pre-farming/herding lineages today belong solely to L0d and L0k.

There is an intriguing possible rider to this conclusion. North Africa has been entirely depopulated and repopulated, at least with respect to mtDNA variation (Pereira et al. 2010), since the time of the Aterian industry, where modern symbolic behavior is attested very early, similar to Southern Africa, and in contrast to Eastern Africa (Barton et al. 2009). We might therefore contemplate a possible North Africa ancestry for L3, with its rapid radiation corresponding to an early range expansion into Eastern Africa. However, any potential dispersal between the Mediterranean and the Horn of Africa around the time of the MIS4/3 transition would face severe environmental difficulties, unlike the “green Sahara” conditions of MIS5 and the early Holocene (Drake et al. 2010). We therefore conclude that an indigenous origin for L3 in Eastern Africa remains by far the most likely scenario.

As Mellars (2006) has argued, the early evidence for symbolically mediated behavior in both North and Southern Africa rules out any simple direct link for the expansion of L3 to (Ambrose 1998; Watson et al. 1997). Evidence of engraved ochre now extends back to at least 100 ka (Henshilwood et al. 2009), *Nassarius* marine shell beads were evidently present across the range of early modern humans from Southern Africa to North Africa and the Levant before 80 ka – possibly tens of thousands of years earlier (Barton et al. 2009; Bouzouggar et al. 2007; d'Errico et al. 2009; Mellars 2006; Vanhaeren et al. 2006) – and evidence for burial ritual is found in early modern humans in the Levant dating to 90–110 ka (Mellars 2006; Shea 2008). Thus, as suggested by Basell (2008) the demographic expansions that led to the first successful dispersal out of Africa seem better explained by the play of palaeoenvironmental forces than by recourse to the advantages of “modernity”.

Major L3 dispersals within Africa also took place during the Holocene, including, but not limited to, the Bantu expansion (Figure 5B). Indeed, the immediate postglacial period may have been at least as significant for reshaping the African genetic landscape. The Lateglacial period of tropical Africa is now thought to have witnessed a mega-drought at 16–17 ka comparable to those prior to 70 ka (Stager et al. 2011). From ~11.5 ka, however, for a few thousand years, the climate was warm and humid, although less so at the northern and southern extremes (Barham and Mitchell 2008; Brooks et al 2005; Kuper and Kröpelin 2006; Weldeab et al. 2007). The Holocene Climatic Optimum allowed population sizes to increase and movement of people probably also took place. We detected increases in effective population sizes at this time both in Eastern Africa and Central/West Africa, which had distinct L3 mtDNA gene pools. Thus the postglacial population expansions were widespread and most likely caused by continent-wide climate change. They mirror the mtDNA expansion signals elsewhere in the world in the early Holocene, such as in Europe (Soares et al. 2010, Atkinson et al. 2008) and Southeast Asia (Soares et al. 2008).

During the Climatic Optimum, genetic exchange occurred between East Africa, Central Africa and the Sahel, with the onset of wetter conditions and the shifting of the desert margins (Brooks et al. 2005; Kuper and Kröpelin 2006). Haplogroup L3f3 is thought to have expanded from Eastern Africa into Chad ~8 ka (Cerný et al. 2009b). The present analysis shows that not only L3f3, but several L3f clades (mainly L3f1), probably also expanded around this time into Central Africa, leading to L3f's widespread distribution in Central Africa, and to its later contribution to the maternal gene pool of Bantu-speaking agriculturists. The North African data also show migrations across the Sahel belt in this time period, perhaps associated with the expansion of herding across the Sahara (Barham and Mitchell 2008), confirming its status as a major corridor for bidirectional migration (Cerný et al. 2007).

The most recent major signal of population expansion in L3 evidently corresponds to the Bantu expansions, which are well-attested by linguistic, archaeological and genetic evidence (Phillipson 2005, Richards et al. 2004). We detected a sharp population size increase ~4 ka ago in the Central African data, the largest witnessed in the L3 data. Gignoux et al. (2011) have also detected population growth for Sub-Saharan Africa in this time frame using a dataset of combined African haplogroups previously assigned to Bantu dispersals (Salas et al. 2002), including L3d and L3e2. The founder analyses from Central Africa into Eastern Africa and Southern Africa showed migrations involving L3b, L3d, L3e and L3f lineages within the last 2–3 ka. The data do not indicate any genetic input involving L3 lineages into Southern Africa before this time; nor is there any distinct signal for more recent immigration

within the last millennium (Huffman 2007). The BSP for Central Africa and the founder analyses to Southern Africa show a strong correspondence between the beginning of the Bantu expansion in Central Africa and the timing of the Bantu dispersals to the south. They thereby corroborate the molecular clocks and dating methodologies that we have employed, and strengthen our conclusions regarding the less well-attested earlier Holocene and late Pleistocene dispersals.

Supplemental data

Supplemental Data include nine figures, four tables, HVSI and HVSII sequences in fasta format and a phylogenetic tree reconstruction.

Acknowledgements

We thank Paul Mellars, Peter Mitchell and Stephen Oppenheimer for critical advice. FCT, the Portuguese Foundation for Science and Technology, supported this work through the research project PTDC/ANT/66275/2006 and the personal grants to VF (SFRH/BD/61342/2009), JBP (SFRH/BD/45657/2008), MDC (SFRH/BD/48372/2008), NS (SFRH/BD/69119/2010) and PS (SFRH/BPD/64233/2009). PS also thanks the de Laszlo Foundation for financial support. The Ministry of Education of the Czech Republic also supported this work through the grant KONTAKT ME 917. IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT.

Web resources

The URLs for data presented herein are as follows:

Calculator for converting ρ values and ML estimates to age estimates,

<http://www.fbs.leeds.ac.uk/staff/profile.php?tag=Richards>

Network 4, <http://www.fluxusengineering.com/sharenet.htm>

mtDNA-GeneSyn, <http://www.ipatimup.pt/downloads/mtDNAGeneSyn.zip>

References

- Ambrose SH. 1998. Chronology of the Later Stone Age and food production in East Africa. *J Arch Sci* **25**: 377-392.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers R, Turnbull D, and Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**: 147.
- Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, and Uerpmann H-P. 2011. The southern route "Out of Africa": Evidence for an early expansion of modern humans into Arabia. *Science* **331**: 453-456.
- Atkinson QD, Gray RD, and Drummond AJ. 2008. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol Biol Evol* **25**: 468-474.
- Atkinson QD, Gray RD, and Drummond AJ. 2009. Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Philos Trans R Soc Lond B Biol Sci* **276**: 367-373
- Bandelt H-J, Alves-Silva J, Guimaraes PEM, Santos MS, Brehm A, Pereira L, Coppa A, Larruga JM, Torroni A, Prata MJ et al. 2001. Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and the Atlantic slave trade. *Ann Hum Genet* **65**: 549-563.
- Bandelt H-J, Forster P, Sykes BC, and Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141**: 743-753.
- Bar-Yosef O. 1992. The role of western Asia in modern human origins. *Philos Trans R Soc Lond B Biol Sci* **337**: 193-200.
- Bar-Yosef O. 1995. The role of climate in the interpretation of human movements and cultural transformations. In *Paleoclimate and evolution, with emphasis on human origins* (eds. ES Vrba GH Denton TC Partridge, and LH Bickele), pp. 507-523. Yale University Press, New Haven.
- Barham L and Mitchell P. 2008. *The first Africans: African archaeology from the earliest toolmakers to most recent foragers* Cambridge.
- Barker G, Barton H, Bird M, Daly P, Datan I, Dykes A, Farr L, Gilbertson D, Harrisson B, Hunt C et al. 2007. The "human revolution" in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J Hum Evol* **52**: 243-261.
- Barton RNE, Bouzouggar A, Collcutt SN, Schwenninger J-L, and Clark-Balzan L. 2009. OSL dating of the Aterian levels at Dar es-Soltan I (Rabat, Morocco) and implications for the dispersal of modern Homo sapiens. *Quaternary Sci Rev* **28**: 1914-1931.
- Basell, LS. 2008. Middle Stone Age (MSA) site distributions in eastern Africa and their relationship to Quaternary environmental change, refugia and the evolution of Homo sapiens. *Quaternary Sci Rev* **27**: 2484-2498.
- Behar DM, Vilems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D et al. 2008. The dawn of human matrilineal diversity. *Am J Hum Genet* **82**: 1130-1140.
- Beyin A. 2006. The Bab al Mandab vs the Nile-Levant: An appraisal of the two dispersal routes for early modern humans out of Africa. *Afr Archaeol Rev* **23**: 5-30.
- Bouzouggar A, Barton N, Vanhaeren M, d'Errico F, Collcutt S, Higham T, Hodge E, Parfitt S, Rhodes E, Schwenninger JL et al. 2007. 82,000-year-old shell beads from North Africa and implications for the origins of modern human behavior. *Proc Natl Acad Sci USA* **104**: 9964-9969.

- Brooks N, Chiapello I, Lernia SD, Drake N, Legrand M, Moulin C, and Prospero J. 2005. The climate-environment-society nexus in the Sahara from prehistoric times to the present day. *J N Afr Stud* **10**: 253-292.
- Bulbeck D. 2007. Where river meets sea: A parsimonious model for *Homo sapiens* colonization of the Indian Ocean rim and Sahul. *Curr Anthropol* **48**: 315-321.
- Campbell MC, Tishkoff SA. 2011. The evolution of human genetic and phenotypic variation in Africa. *Curr Biol* **20**: R166-173
- Cerný V, Pereira L, Kujanová M, Vasíková A, Hájek M, Morris M, Mulligan CJ. 2009a. Out of Arabia-the settlement of island Soqatra as revealed by mitochondrial and Y chromosome genetic diversity. *Am J Phys Anthropol* **138**: 439-447
- Cerný V, Fernandes V, Costa MD, Hájek M, Mulligan CJ, and Pereira L. 2009b. 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.
- Cerný V, Salas A, Hájek M, Zaloudková M, and 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**: 433-452.
- Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, Frigi S, Yacoubi-Loueslati B, Amor MB, Slama A, Amorim A et al. 2009. Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. *Am J Phys Anthropol* **139**: 253-260.
- Cohen AS, Stone JR, Beuning KR, Park LE, Reinthal PN, Dettman D, Scholz CA, Johnson TC, King JW, Talbot MR et al. 2007. Ecological consequences of early Late Pleistocene mega-droughts in tropical Africa. *Proc Natl Acad Sci USA* **104**: 16422-16427.
- Compton JS. 2011. Pleistocene sea-level fluctuations and human evolution on the southerncoastal plain of South Africa. *Quaternary Sci Rev* **30**: 506-527
- d'Errico F, Vanhaeren M, Barton N, Bouzouggar A, Mienis H, Richter D, Hublin J-J, McPherron SP, and Lozouet P. 2009. Additional evidence on the use of personal ornaments in the Middle Paleolithic of North Africa. *Proc Natl Acad Sci USA* **106**: 16051-16056
- Derricourt R. 2005. Getting "Out of Africa": Sea crossings, land crossings and culture in the hominin migrations. *J World Prehist* **19**: 119-132.
- Drake NA, Blench RM, Armitage SJ, Bristow CS, White KH. 2011. Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proc Natl Acad Sci USA* **108**: 458-462
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4** : e88
- Drummond A and Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**: 214.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**: 1185-1192.
- Endicott P and Ho SYW. 2008. A bayesian evaluation of human mitochondrial substitution rates. *Am J Hum Genet* **82**: 895-902.
- Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Silva WA, Zago MA, Ribeiro-dos-Santos AK, Santos SEB, Petzl-Erler ML, Bonatto SL. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* **82**: 583-592
- Field JS and Lahr MM. 2005. Assessment of the southern dispersal: GIS-based analyses of potential routes at oxygen isotopic stage. *J World Prehist* **19**: 1-45.

- Forster P. 2004. Ice ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc Lond B Biol Sci* **359**: 255-264.
- Forster P, Harding R, Torroni A, and Bandelt H-J. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* **59**: 935-945.
- Gignoux CR, Henn BM, Mountain JL. 2011. Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci USA* **108**: 6044-6049
- Grün R, Stringer C, McDermott F, Nathan R, Porat N, Robertson S, Taylor L, Mortimer G, Eggins S, and McCulloch M. 2005. U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *J Hum Evol* **49**: 316-334.
- Harich N, Costa MD, Fernandes V, Kandil M, Pereira JB, Silva NM, and Pereira L. 2010. The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evol Biol* **10**: 138.
- Haslam M, Clarkson C, Petraglia M, Korisettar R, Bora J, Boivin N, Ditchfield P, Jones S, and Mackay A. 2010. Indian lithic technology prior to the 74,000 BP Toba super-eruption: searching for an early modern human signature. In *Upper Palaeolithic Revolution in Global Perspective: Essays in honour of Paul Mellars* (eds. V Boyle C Gamble, and O Bar-Yosef), pp. 73-84. McDonald Institute for Archaeological Research, Cambridge.
- Henshilwood CS, d'Errico F, and Watts I. 2009. Engraved ochres from the Middle Stone Age levels at Blombos Cave, South Africa. *J Hum Evol* **57**: 27-47.
- Henshilwood CS, d'Errico F, Yates R, Jacobs Z, Tribolo C, Duller GAT, Mercier N, Sealy JC, Valladas H, Watts I et al. 2002. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science* **295**: 1278-1280.
- Huffman TN. 2007. *Handbook to the Iron Age: The archaeology of pre-colonial farming societies in Southern Africa*. University of KwaZulu-Natal Press, KwaZulu-Natal.
- Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* **72**: 313-332.
- Klein RG. 1992. The archeology of modern human origins. *Evol Anthropol Issues News Rev* **1**: 5-14.
- Kuper R and Kröpelin S. 2006. Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution. *Science* **313**: 803-807.
- Lawler A. 2011. Did modern humans travel out of Africa via Arabia? *Science* **331**: 387.
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**: 1034-1036.
- Mellars P. 2006. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci USA* **103**: 9381-9386.
- Newman JL. 1995. *The peopling of Africa: a geographic interpretation*. Yale University Press, New Haven.
- Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon J-M et al. 2006. The mtDNA legacy of the Levantine Early Upper Palaeolithic in Africa. *Science* **314**: 1767-1770.
- Oppenheimer S. 2003. *Out of Eden*. Constable and Robinson Ltd, London.
- Ottoni C, Primativo G, Hooshiar KB, Achilli A, Martínez-Labarga C, Biondi G, Torroni A, and Rickards O. 2010. Mitochondrial haplogroup H1 in North Africa: an early Holocene arrival from Iberia. *Plos One* **5**: e13378.
- Pereira L, Freitas F, Fernandes V, Pereira JB, Costa MD, Costa S, Máximo V, Macaulay V, Rocha R, and Samuels DC. 2009. The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet* **84**: 628-640.

- Pereira L, Goncalves J, Franco-Duarte R, Silva J, Rocha T, Arnold C, Richards M, and Macaulay V. 2007. No evidence for an mtDNA role in sperm motility: data from complete sequencing of asthenozoospermic males. *Mol Biol Evol* **24**: 868-874.
- Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, and 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 L, Prata MJ, and Amorim A. 2000. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* **64**: 491-506.
- Pereira L, Silva NM, Franco-Duarte R, Fernandes V, Pereira JB, Costa MD, Martins H, Soares P, Behar DM, Richards MB et al. 2010. Population expansion in the North African Late Pleistocene signalled by mitochondrial DNA haplogroup U6. *BMC Evol Biol* **10**: 390.
- Petraglia M, Korisettar R, Boivin N, Clarkson C, Ditchfield P, Jones S, Koshy J, Lahr MM, Oppenheimer C, Pyle D et al. 2007. Middle Palaeolithic assemblages from the Indian Subcontinent before and after the Toba super-eruption. *Science* **317**: 114-116.
- Phillipson DW. 2005. *African archaeology*. Cambridge University Press, Cambridge.
- Richards M, Bandelt HJ, Kivisild T, and Oppenheimer S. 2006. A model for the dispersal of modern humans out of Africa. In *Mitochondrial DNA and the evolution of Homo sapiens*. (eds. HJ Bandelt V Macaulay, and M Richards), pp. 225–265. Springer-Verlag, Berlin.
- Richards M, Macaulay V, Hill C, Carracedo A, Salas A. 2004. The archaeogenetics of the Bantu dispersals. In: *Traces of Ancestry: Studies in honour of Colin Renfrew* (eds. M Jones) pp. 75-88. McDonald Institute for Archaeological Research, Cambridge.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**: 1251-1276.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, and Nørby SS. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* **67**: 718-726.
- Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sanchez-Diz P, Macaulay V, and Carracedo A. 2002. The making of the African mtDNA landscape. *Am J Hum Genet* **71**: 1082-1111.
- Scheinfeldt LB, Soi S, Tishkoff SA. 2010. Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proc Natl Acad Sci USA* **107** Suppl 2:8931-8938.
- Scholz CA, Johnson TC, Cohen AS, King JW, Peck JA, Overpeck JT, Talbot MR, Brown ET, Kalindekafé L, Amoako PY et al. 2007. East African megadroughts between 135 and 75 thousands years ago and bearing on early-modern human origins. *Proc Natl Acad Sci USA* **104**: 16416-16421.
- Shea JJ. 2008. Transitions or Turnovers? Climatically-Forced Extinctions of *Homo sapiens* and Neandertals in the East Mediterranean Levant. *Quaternary Sci Rev* **27**: 2253-2270
- Shi W, Ayub Q, Vermeulen M, Shao RG, Zuniga S, van der Gaag K, de Knijff P, Kayser M, Xue Y, and Tyler-Smith C. 2010. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol Biol Evol* **27**: 385-393.
- Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, Torroni A, and Richards MB. 2010. The Archaeogenetics of Europe. *Curr Biol* **20**: R174-R183.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, and Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* **84**: 740-759.

- Soares P, Rito T, Trejaut J, Mormina M, Hill C, Tinkler-Hundal E, Braid M, Clarke DJ, Loo J-H, Thomson N et al. 2011. Ancient Voyaging and Polynesian Origins. *Am J Hum Genet* **88**: 239-247.
- Soares P, Trejaut JA, Loo J-H, Hill C, Mormina M, Lee C-L, Chen Y-M, Hudjashov G, Forster P, Macaulay V et al. 2008. Climate change and post-glacial human dispersals in Southeast Asia. *Mol Biol Evol* **25**: 1209-1218.
- Stager C, Ryves DB, Chase BM, Pausata FSR. 2011. Catastrophic drought in the Afro-Asian monsoon region during Heinrich Event 1. *Science* **331**: 1299-1302
- Texier PJ, Porraz G, Parkington J, Rigaud JP, Poggenpoel C, Miller C, Tribolo C, Cartwright C, Coudenneau A, Klein R, Steele T, Verna C. 2010. A Howiesons Poort tradition of engraving ostrich eggshell containers dated to 60,000 years ago at Diepkloof Rock Shelter, South Africa. *Proc Natl Acad Sci USA* **107**: 6180-6185
- Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, and Singh L. 2005. Reconstructing the origin of Andaman Islanders. *Science* **308**: 996.
- Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O et al. 2009. The genetic structure and history of Africans and African Americans. *Science* **324**: 1035-1044.
- Torrioni A, Achilli A, Macaulay V, Richards M, and Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet* **22**: 339-334
- Turney CSM, Kershaw AP, Moss P, Bird MI, Fifield LK, Cresswell RG, Santos GM, Tada MLD, Hausladen PA, and Zhou Y. 2001. Redating the onset of burning at Lynch's Crater (North Queensland): implications for human settlement in Australia. *J Quaternary Sci* **16**: 767-771.
- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**: 386-394
- Vanhaeren M, d'Errico F, Stringer C, James SL, Toadd JA, and Mienis HK. 2006. Middle Paleolithic shell beads in Israel and Algeria. *Science* **312**: 1785-1788.
- Watson E, Forster P, Richards M, and Bandelt H-J. 1997. Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* **61**: 691-704.
- Weldeab S, Lea DW, Schneider RR, and Andersen N. 2007. 155,000 years of West African monsoon and ocean thermal evolution. *Science* **316**: 1303-1307.
- Willoughby PR. 2007. *The evolution of modern humans in Africa - A comprehensive guide*. AltaMira Press, Lanham.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**: 555-556.
- Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time *Am J Hum Genet* **74**:50-61

Figure legends

Figure 1. Schematic tree of haplogroup L3. Age estimates are respectively ρ estimates from the complete mtDNA genome (black font), ρ estimates using the synonymous clock (grey font) and maximum likelihood estimates using the complete mtDNA genome (black font underlined). Colour scheme for each clade indicates the probable geographic origin. Haplogroups M and N are indicated, but no age estimate is provided since they were not calculated in this work. MSA refers to Middle Stone Age and LSA to Late Stone Age.

Figure 2. Bayesian Skyline Plot (BSP) indicating the median of the hypothetical effective population size through time based on data from the entire L3 haplogroup, except for haplogroups M and N, assuming a generation time of 25 years. Maximum time (x-axis) corresponds to the median posterior estimate of the genealogy root-height.

Figure 3. Frequency maps based on HVS-I data for haplogroups L3a, L3i, L3h and L3x combined (A), L3f (B), L3e (C), L3b (D) and L3d (E).

Figure 4. Probabilistic distribution of founder clusters across migration times scanned at 200 years intervals from 0–100 ka, using the *f_I* criterion, for Eastern Africa (A), Central Africa (B), North Africa (C) and Southern Africa (D; overall dataset in black and Eastern Bantu route data in grey).

Figure 5. Outline of the main migrations detected in this work during the Pleistocene (A) and the Holocene (B).

Table 1. Peaks of rate of population size change through time as obtained from the BSPs for the general and geographically defined data and periods of time where the rate of population size increase was of at least one individual per 100 individuals in a period of 100 years. Increment ratio corresponds to the number of times the effective population size increase during this period.

Dataset	Peak	Range of increment	Increment ratio
Overall	40.3	38.5-43.3	2.89
	10.9	9.1-12.4	4.59
	3.3	1.7-4.3	9.52
East Africa	36.8	31.7-41.6	4.72
	9.2	6.7-11.3	5.89
Central Africa	9.9	8.2-11.5	2.59
	2.6	1.0-3.7	7.59
North Africa	10.0	7.3-12.4	8.55
	35.7	32.5-38.5	2.17

Table 2. Percentage of lineages that entered East, Central, North and South Africa assuming specific migration times. Migration times were obtained from a previous scan using increments of 200-years intervals resulting in probable peaks of migration. A migration time of 0.1 ka was included with each analysis to account for historical gene flow, with the exception of the founder analysis into North Africa where the 0.6 ka value was already close to this value.

Geography	Founder Analysis	
	Migration time (ka)	% of L3 lineages (SE)
East Africa	58.8	74.0 (0.5)
	1.8	20.1 (2.6)
	0.1	5.9 (2.5)
Central Africa	42.4	75.0 (2.7)
	9.2	24.1 (2.8)
	0.1	0.9 (0.2)
North Africa	35.0	7.4 (2.7)
	6.6	67.0 (4.0)
	0.6	25.7 (3.1)
South Africa (whole)	3.2	86.7 (4.3)
	0.1	13.3 (4.3)
South Africa (southern)	1.8	83.4 (3.7)
	0.1	16.6 (3.7)

Figure 1.

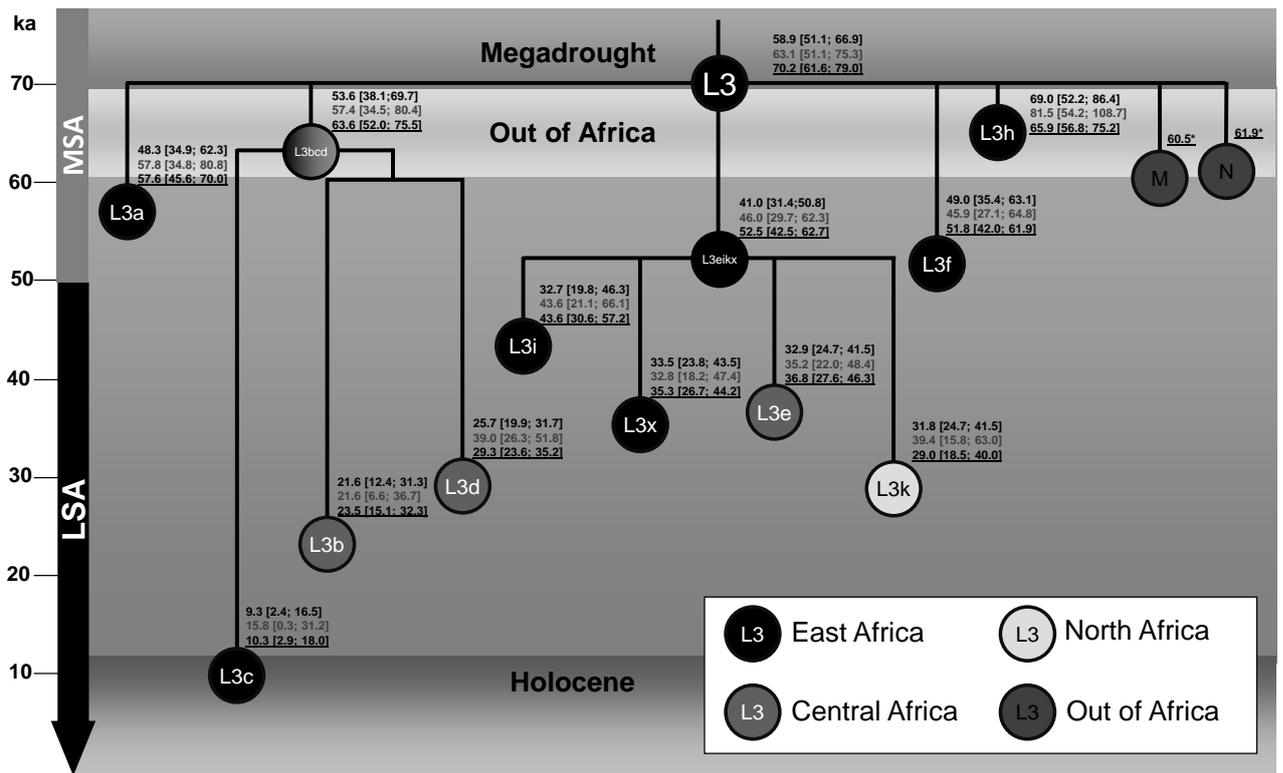


Figure 2

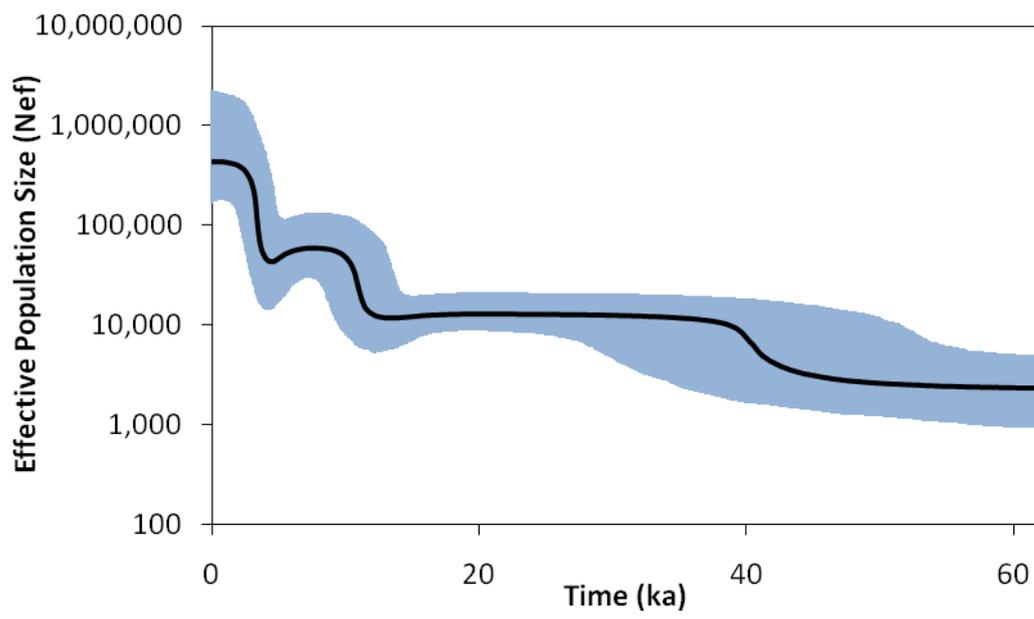


Figure 3.

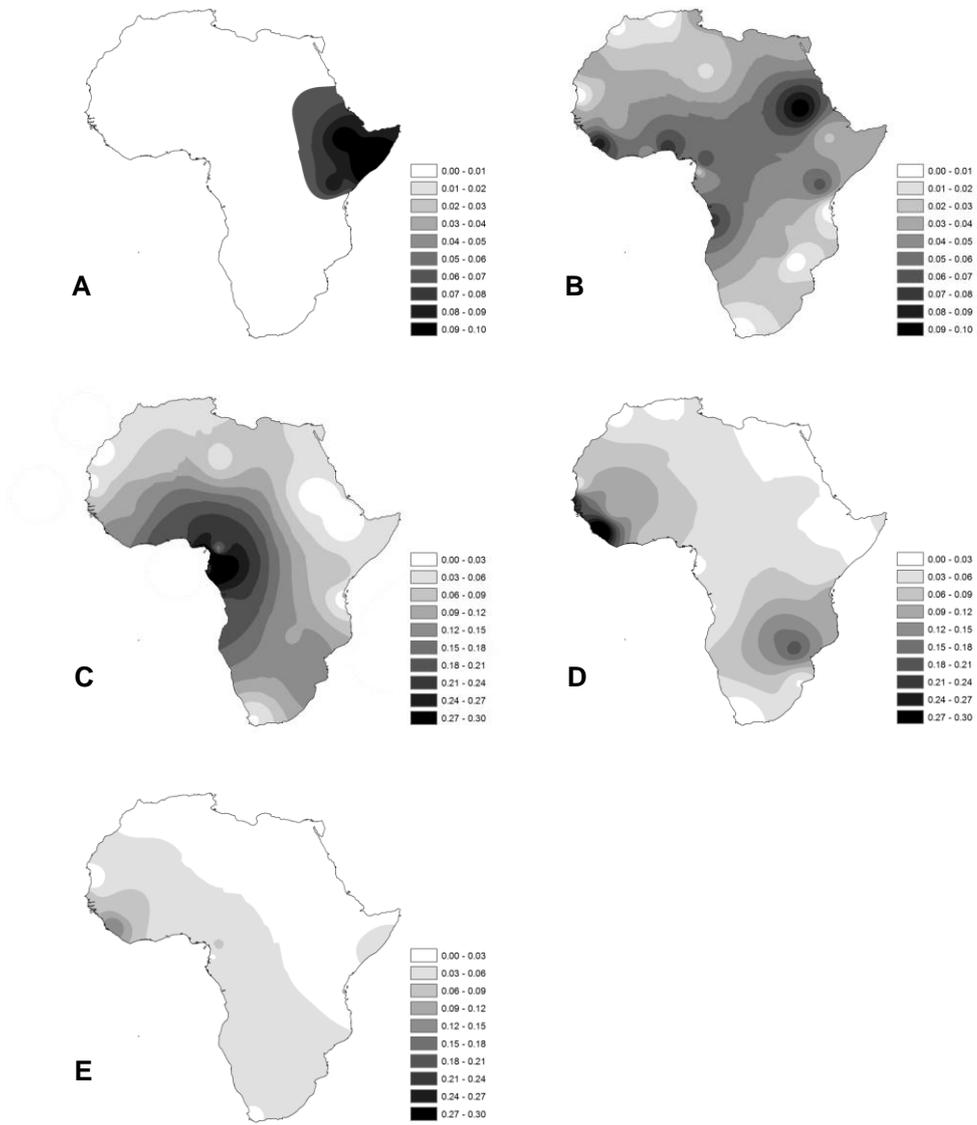


Figure 4.

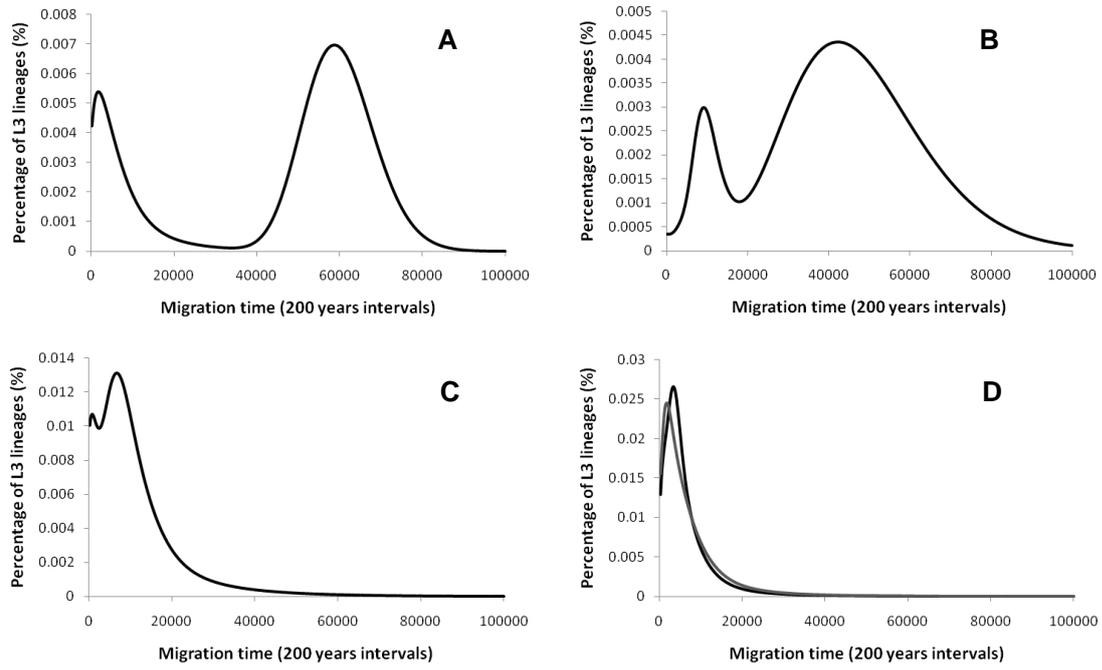


Figure 5.

