Atkinson, Q. D. Gray, R.D. and Drummond, A. J. (2009). Bayesian coalescent inference of major human mtDNA haplogroup expansions in Africa. Proc. Roy. Soc. Lon. B. 276: 367-373.

Bayesian coalescent inference of major human mtDNA haplogroup expansions in Africa

Quentin D. Atkinson[1*], Russell D. Gray[2] and Alexei J. Drummond[3]

1 Institute of Cognitive and Evolutionary Anthropology, University of Oxford, 64 Banbury Rd, Oxford, OX2 6PN, UNITED KINGDOM

2 Department of Psychology, University of Auckland, Private Bag 92019, Auckland 1142, NEW ZEALAND

3 Bioinformatics Institute and Department of Computer Science, University of Auckland, Private Bag 92019, Auckland, NEW ZEALAND

* Corresponding author. Email: quentin.atkinson@anthro.ox.ac.uk

Summary

Past population size can be estimated from modern genetic diversity using coalescent theory. Estimates of ancestral human population dynamics in Africa can tell us about the timing and nature of our first steps towards colonizing the globe. Here we combine Bayesian coalescent inference with a dataset of 224 complete human mtDNA sequences to estimate effective population size through time for each of the four major African mtDNA haplogroups (L0-L3). We find evidence of three distinct demographic histories underlying the four haplogroups. Haplogroups L0 and L1 both show slow, steady exponential growth from 156-213 kya. In contrast, haplogroups L2 and L3 show evidence of substantial growth beginning 12-20kya and 61-86 kya respectively. These later expansions may be associated with contemporaneous environmental and/or cultural changes. The timing of the L3 expansion - 8-12ky prior to the emergence of the first non-African mtDNA lineages - together with high L3

diversity in east Africa, strongly supports the proposal that the human exodus from Africa and subsequent colonization of the globe was prefaced by a major expansion within Africa, perhaps driven by some form of cultural innovation.

Key index words or phrases - mitochondrial DNA, Africa, phylogenetics, Bayesian skyline plot, coalescent theory

Introduction

Phylogenetic analyses of human genetic diversity are revealing an increasingly detailed picture of the human colonization of the globe (Forster & Matsumura 2005; Macaulay *et al.* 2005; Thangaraj *et al.* 2005; Olivieri *et al.* 2006; Witherspoon *et al.* 2006; Underhill & Kivisild 2007; Atkinson *et al.* 2008). A late Pleistocene African origin of modern humans is now widely accepted, and is supported by mitochondrial DNA (mtDNA) (Cann *et al.* 1987; Vigilant *et al.* 1991; Watson *et al.* 1997; Ingman *et al.* 2000), Y-chromosome (Underhill *et al.* 2000; Underhill & Kivisild 2007) and autosomal (Gonser *et al.* 2000; Zhivotovsky *et al.* 2000; Alonso & Armour 2001; Rogers, 2001; Marth *et al.* 2004; Witherspoon *et al.* 2006) evidence. More recent, mtDNA-based work suggests a human expansion from Africa along the Indian Ocean coast 50-70kyrs ago (Forster & Matsumura, 2005; Macaulay *et al.* 2005; Thangaraj *et al.* 2005) linked with substantial population growth in South Asia (Atkinson *et al.* 2008), followed by later expansions into Sahul (Friedlaender *et al.* 2005; Atkinson *et al.* 2008), the Levant and North Africa (Olivieri *et al.* 2006), Europe (Atkinson *et al.* 2008) and the New World (Forster *et al.* 1996; Atkinson *et al.* 2008). Within Africa, however, the nature of population expansions before, during and after the global diaspora remains unclear.

Together with evidence from archaeology and other genetic loci, mtDNA diversity can help elucidate human prehistory in Africa, just as it has on a global scale. The absence of recombination, combined with a high copy number and fast rates of genetic change, make mtDNA well-suited to phylogenetic analysis over the timescale of modern humans. Inferences must be made with the caveat that, by virtue of being inherited from mother to child, mtDNA can only capture the history of the maternal

lineage. In addition, as with any single-locus analysis, parameter estimates are only valid to the extent that the locus reflects population demographic history and not the influence of selection or changes in population structure. However, by combining mtDNA-based findings with evidence from other genetic loci, as well as archaeology and linguistics, we can resolve an increasingly detailed picture of human prehistory.

The time to the most recent common ancestor (TMRCA) of the human mtDNA tree is dated to between approximately 150kya and 250kya (Mishmar *et al.* 2003; Macaulay *et al.* 2005; Atkinson *et al.* 2008). Lineages indigenous to Africa occur at the base of the mtDNA tree and are usually divided into 3 or 4 main haplogroups – L0 (formerly L1a, L1d, L1f and L1k), L1, L2, and L3 (see Figure 1). Four other haplogroups are sometimes identified (L4-L7) but these are relatively rare (Gonder *et al.* 2007). Haplogroup L0 is found in southern Africa and east/southeast Africa (Salas *et al.* 2002; Gonder *et al.* 2007). Its oldest branches (L0d and L0k) occur at high frequencies only amongst the Khoisan hunter-gatherers of South Africa (Watson *et al.* 1997; Salas *et al.* 2002). Haplogroup L1 is found at moderately high frequencies in western and central Sub-Saharan Africa, including in the pygmy populations of the central equatorial forest region (Salas *et al.* 2002). Haplogroup L2 is common in western and south-eastern sub-Saharan Africa and is the most numerous and widespread of the 4 major haplogroups, accounting for as much as 25% of indigenous haplotype variation (Salas *et al.* 2002). The youngest of the major haplogroups, L3, is most common in western and east/southeast sub-Saharan Africa, particularly amongst speakers of the Bantu language family, and is thought to have originated in east Africa, where it accounts for half of all types (Salas *et al.* 2002). Two other

haplogroups descended from L3 (M and N) are the only mtDNA lineages observed outside Africa.

The different temporal and geographic distributions of these four haplogroups make them potentially useful as markers of broad population demographic processes within Africa over the last 150,000 years. The oldest lineages, haplogroups L0 and L1, occur at higher frequencies among hunter-gatherers and may provide a window into our earliest history. For example, the occurrence of certain L0 haplotypes amongst the Khoisan has recently been shown to support a split between the ancestral Khoisan population and other lineages between 90kya and 150kya (Behar *et al.* 2008). The younger L2 and L3 haplogroups are much more widespread and common, but it is unclear how they achieved their current distribution and why the L3 haplogroup was the only lineage to be carried out of Africa.

Information about ancestral population size and growth parameters can be inferred from extant genetic variation using coalescent theory (Kingman 1982; Hudson 1990; Griffiths & Tavaré 1994). Early coalescent estimates of past African population demographics were based on the distribution of pair-wise sequence differences in RFLP and D-loop data, assumed a single exponentially growing population and included a large degree of statistical uncertainty (Harpending *et al.* 1993; Sherry *et al.* 1994; Harpending & Rogers 2000). More recent work using complete mtDNA sequences has focussed on the geographic and phylogenetic relationships between specific haplotypes and their TMRCAs but has not explicitly modelled population size through time (Watson *et al.* 1997; Salas *et al.* 2002; Bandelt *et al.* 2001; Gonder *et al.* 2007). Watson *et al.* (1997) were able to use median networks to identify star-

like patterns in L2 and L3 RFLP and D-loop sequence diversity, which they suggest reflect an expansion 80-60kya, possibly linked to the subsequent human expansion from Africa. However, without model-based coalescent inference it is difficult to evaluate hypotheses about the dispersal of the different haplogroups. Inferred lineage coalescence events constitute a single realization of a stochastic process, meaning there is no simple correspondence between lineage TMRCAs and population expansions.

Improvements in the available data and inference methods mean that it is now possible to estimate population demographic parameters through time with increased accuracy, without assuming a simple parametric growth curve and with an explicit framework for quantifying uncertainty in population parameter estimates (Shapiro *et al.* 2004; Atkinson *et al.* 2008). Given a sample of sequences from a population, the Bayesian Skyline plot (BSP) (Drummond & Rambaut 2003; Drummond *et al.* 2005) uses Bayesian coalescent inference together with a Markov Chain Monte Carlo (MCMC) (Metropolis *et al.* 1953; Hastings 1970) sampling algorithm to provide a credibility interval for effective population size through time that incorporates uncertainty in the substitution model parameters, underlying genealogy and coalescent times. Unlike previous approaches to estimating population size, the BSP does not require a pre-specified parametric growth model, and because the method is applied directly to a set of sequence data rather than to pair-wise sequence differences, it avoids the loss of information that is associated with distance-based methods (Steel *et al.* 1988). Elsewhere (Atkinson *et al.* 2008), we have used BSPs together with a large dataset of complete human mtDNA sequences to estimate the relative size and timing of regional human population expansions around the globe. We showed that

coalescent estimates of present-day effective population sizes from complete mtDNA sequence diversity were a good predictor of relative contemporary population sizes inferred from historical and anthropological data. Estimates of ancestral effective population size in Africa revealed slow, relatively constant growth, but because the continent was treated as a single population, we would not expect to detect the expansion of any sub-group of the population that did not substantially change the overall population size.

Here we use BSPs together with a dataset of African complete mtDNA sequences to estimate the relative population size of the four major mtDNA haplogroups through time. These plots do not equate to enduring physically separated populations, which are difficult to identify in Africa, but rather represent the expansion of four separate mitochondrial lineages within the African meta-population. It is expected that population sizes will be strongly correlated further back in time, reflecting the small founding population that the lineages arose from. By plotting the lineages separately we can answer questions about the timing of later expansions within Africa that may have been linked to populations carrying the different haplogroups. For example, do haplogroups L0 and L1 show significantly different demographic histories, perhaps suggesting population divisions amongst the earliest humans? When did L2, Africa's most common haplogroup, begin to spread and how quickly did this occur? And, perhaps most intriguing, when did the L3 lineage begin to spread in Africa and what can this tell us about our subsequent expansion from our homeland?

Materials and Methods

Sequence data

We compiled a dataset of 224 complete African mtDNA sequences from those reported in the literature (Ingman *et al.* 2000; Maca-Meyer *et al.* 2001; Mishmar *et al.* 2003; Macaulay *et al.* 2005; Kivisild *et al.* 2006; Torroni *et al.* 2006; Gonder *et al.* 2007)(supplementary table S1, Electronic Supplementary Material). Only sequences identified as L haplotypes were sampled, and these were classified as either L0 (n=60), L1 (n=41), L2 (n=43) or L3 (n=80). We included L4 (also known as L3g) as part of L3 and ignored the rarer haplotypes L5-L7. Excluding L4 from the analysis did not noticeably affect the results for L3 (supplementary figure S1, Electronic Supplementary Material). Sequences were machine aligned in *ClustalX* (Thompson *et al.* 1997) using reference sequence J01415.1. The coalescent inference method we used to analyse the data assumes that sequences are sampled at random from the population of interest. Here, because we are treating each haplogroup as a separate population, we avoided sampling sequences from studies that focussed on particular sub-types within an African haplogroup. Of the remaining studies, designed to sample African mtDNA diversity, complete sequencing may be non-random due to patchy sampling and the desire to sequence representatives of rarer variants. We do not expect this to significantly affect our findings, because we are combining sequences from a number of independent studies and there is no reason to think the sampling procedure will have systematically biased the representation of haplogroups across all of these studies. Repeating the analyses using subsets of the full dataset, in which sequences from the larger contributing studies (Kivisild *et al.* 2006; Gonder *et al.* 2007) were removed, did not affect our results, producing the same patterns that we report below.

Bayesian Skyline Plots

For each of the four L haplogroups, BSPs (Drummond *et al.* 2005) of effective population size through time were constructed using a MCMC (Metropolis *et al.* 1953; Hastings 1970) sampling algorithm, as implemented in *BEAST* (*version 1.4*)(Drummond & Rambaut 2007). We also constructed BSPs of sets of 100 sequences randomly sampled from the entire 224-sequence dataset to estimate total African population size through time. The underlying population size function of the BSP can be fitted using either a piecewise constant or a piecewise linear function of population size change. Here we report results based on a piecewise linear model made up of 10 control points. The qualitative results were not affected by the number of control points fitted or by using a piecewise constant function of effective population size. Effective population size is a compound population genetics parameter generally considered to be linearly proportional to census population size – here the population of breeding females carrying haplotypes from each haplogroup. Other factors, such as local extinction with recolonization, selection and various forms of non-random mating (Wakeley 2000), can also influence effective population size estimates. As we discuss below, selective sweeps are unlikely to account for the patterns we observe and coalescent inference of population size from mtDNA diversity has been shown to be a good predictor of census population size in humans (Atkinson *et al.* 2008).

To infer the ancestral gene trees for each haplogroup we used a general time-reversible (GTR) substitution model with site-specific rates for 1st, 2nd and 3rd codon

positions. This model has been shown to fit the data better than the commonly used GTR + Γ + I model (Shapiro *et al.* 2006; Atkinson *et al.* 2008). Each MCMC sample was based on a run of 40,000,000 generations, sampled every 4,000, with the first 4,000,000 generations discarded as burn-in. Examination of autocorrelation times of the MCMC plots indicated runs had converged to the posterior distribution and provided an adequate sample. All runs had an effective sample size of at least 1000 for the parameters of interest.

In order to assign a time-scale to the population size estimates, the rate of molecular evolution must be calibrated. Evidence for the existence of trends in observed substitution rates through time (Ho *et al.* 2005) makes an internal rate calibration (of similar time depth to the period of interest in Africa) preferable to the commonly applied human-chimp calibration. Following previous work analysing a global sample of human mtDNA diversity (Atkinson *et al.* 2008), rates were fixed to 1.691e-8 substitutions/site/yr. In order to separate uncertainty in population size estimates from rate calibration error, uncertainty in the calibration time itself was not included in the analyses reported here. This allows for a more accurate comparison of the relative timing of growth patterns between the haplogroups. The rate calibration is based on a 45ky age for haplogroup Q in New Guinea and earliest archaeological evidence of human habitation on the island is 42-45kya (O'Connell & Allen 2004). Haplogroup Q occurs at high frequencies only in New Guinea (a closely related haplogroup, Q2, found in parts of Melanesia, also occurs at low frequencies in Micronesia and Polynesia, most likely due to more recent gene flow [Friedlaender *et al.* 2005]). Ho & Endicott (2008) have suggested an alternative calibration linking haplogroup P to the colonization of Sahul, which yields slightly faster rates if the same colonization time

is assumed for Sahul. We note, however, that absolute time estimates scale in proportion to the chosen calibration age, such that if haplogroup Q is assumed to be 40ky old, all inferred dates would simply decrease by approximately 11%.

Results and Discussion

Figure 2 shows BSPs of effective population size through time for the total African population and for each of the four major haplogroups. The bold line represents the median population size estimate from the Bayesian posterior distribution. Gray lines delimit the 95% highest posterior density (HPD) boundaries. Consistent with our previous findings using a smaller African sample (Atkinson *et al.* 2008), when considered as a single population, total African effective population size estimates (Figure 2a) show gradual, roughly exponential growth from a TMRCA 150-201 kya (95% HPD) – with perhaps a slight increase in growth rate ~70kya. As expected, the same gradual growth pattern is revealed when population size estimates are summed across the separate lineage analyses (dotted line, Figure 2a). This pattern contrasts with evidence for rapid growth outside Africa (Atkinson *et al.* 2008) and is consistent with findings from independent genetic loci, including the X- and Y-chromosome (Garrigan *et al.* 2007), and microsatelite and SNPs data (Gonser *et al.* 2000; Zhivotovsky *et al.* 2000; Alonso & Armour 2001; Rogers 2001; Marth *et al.* 2004). This broad agreement across multiple loci makes it likely that the growth signals we infer reflect population expansions rather than multiple simultaneous selective sweeps.

However, the separate BSPs for the four major haplogroups (Figure 2b-e) reveal very different patterns of growth to the overall picture, particularly for haplogroups L2 and L3. Examination of the plots indicates that growth signals in haplogroups L0 (Figure 2b) and L1 (Figure 2c) are not significantly different from each other, showing substantial overlap in HPD distributions through time. However, HPD distributions for haplogroups L2 (Figure 2d) and L3 (Figure 2e) are significantly different from each other and from haplogroups L0 and L1. These differences may reflect changes in the size or structure of the populations carrying these haplotypes or could indicate the influence of selection pressures causing advantageous new variants to sweep through the population. We discuss the implications of the different growth profiles below.

Haplogroups L0 and L1 (Figure 2b and 1c respectively) show slow constant growth over the last 100 to 200 kya (TMRCAs: L0, 124-172 kya; L1, 87-139 kya; L0 and L1 combined, 156-213 kya; 95% HPDs). Estimates of effective population size do not differ significantly throughout the history of the two lineages. We would expect to see this pattern if both lineages formed part of an early panmictic African population and contributed relatively equally to the haplotype diversity present in Africa today. However, recent evidence indicates some population structure within Africa from an early stage. Analysis of genetic data from four African populations using a coalescent based model of population divergence found support for population structure in Africa from >50kya when applied to mtDNA, although not when applied to Y-chromosome data (Garrigan *et al.* 2007). Others have argued that the distribution of ancient mtDNA L0d and L0k lineages amongst the hunter-gatherer speakers of the African ‘click’ languages also supports population structure deep in the mtDNA tree (Knight *et al.* 2003; Tishkoff *et al.* 2007; Behar *et al.* 2008). Our findings show that,

whether such deep population structure exists or not, it did not produce substantially different population growth profiles between the most ancient lineages. Repeating the BSP analysis on only L0d and L0k lineages also did not produce significantly different growth profile from the L0 BSP (supplementary figure S2, Electronic Supplementary Material). This makes it less likely that such putative deep divergence events were associated with a substantial change in available territory or mode of living. Effective population size plots for the two haplogroups do appear to be on different trajectories over the last 15k years, although they do not differ significantly. L0 shows an increasing trend whilst L1 is decreasing. Inspection of the underlying phylogeny indicates that the increase in L0 is likely a result of a rapid emergence of lineages within the L0a haplogroup at this time. L0a has been proposed as a marker of the Bantu expansion 3,000 to 4,000 years ago (Bandelt *et al.* 1995; Chen *et al.* 1995). The time scale we reconstruct would suggest it began to expand earlier, perhaps linked to the expansion of L2a, discussed below.

For L2 we infer a TMRCA of 73-127 kya (95% HPD). This range is consistent with previous age estimates for the haplogroup based on D-loop and complete sequence data and has lead to the suggestion that L2 may have been involved in population expansions associated with the later African exodus (Watson *et al.* 1997; Salas *et al.* 2002). However, the BSP in Figure 2d shows L2 to have been at relatively low frequency until a period of substantial growth beginning 12-20kya (95% HPD). This is the most pronounced expansion signal we observe amongst the major haplogroups and probably explains the high frequency of L2 lineages across Africa today. The most common haplogroup within L2 is L2a (TMRCA 25-31kya). Repeating the analysis on just the L2a lineages confirms that they are the principal source of the L2

expansion signal (supplementary figure S3, Electronic Supplementary Material). Whilst this rapid expansion signal in L2 could have been caused by a selective sweep, the substitutions that define the L2 (and L2a) lineages are synonymous changes, making it unlikely that these were new and advantageous variants capable of quickly displacing existing haplotypes. Like L0a mentioned above, L2a has been linked to the spread of Bantu languages (Bandelt *et al.* 2001). However, as with L0a, the timing of the growth we infer here predates the Bantu expansion. This discrepancy could be explained as a result of the shift in inferred rates of mtDNA evolution at shallow time depths identified by Ho *et al.* (2005), although rates would need to be four to five times faster to bring the start of the growth phase seen in L2a into agreement with the proposed spread of Bantu 3000 to 4000 years ago. Whilst the Bantu expansion undoubtedly played a role in the spread of a number of mtDNA haplotypes, L2 appears to have begun to expand somewhat earlier. Our findings fit with a proposal that L2 lineages (and perhaps the L0a lineages) spread as a result of environmental changes associated with the Last Glacial Maximum (LGM) (Salas *et al.* 2002). Particularly arid conditions at this time are thought to have resulted in the enlargement of the Sahara as well as the conversion of much of the forested area of central Africa to open savannah and woodlands (Adams & Faure, 1997), which may have allowed an expansion of human populations into new territories, particularly in central Africa.

Figure 2e shows the BSP for haplogroup L3 and reveals a marked increase in effective population size from an estimated TMRCA of 61-86 kya (95% HPD). L3 is the only lineage to show such marked growth at this time. Again, the substitutions that define L3 are synonymous changes, hence the inferred increase in L3 frequency is

unlikely to be the result of a selective sweep. Instead, it seems likely that L3 spread due to demographic expansion within Africa. L3 haplotype frequencies are highest in East Africa (Salas *et al.* 2002), the proposed launching point of the human colonization of the globe. The timing of the L3 expansion predates the emergence of the first non-African lineages (haplogroups M and N) by 8-12 ky - based on the same rate calibration used here, we estimate an age of 53-69 kya and 50-64 kya (95% HPDs) for M and N respectively (Atkinson *et al.* 2008). The fact that L3 is the only haplogroup with descendents outside Africa and shows a clear growth signal 8-12ky prior to the emergence of its non-African descendents, strongly suggests that L3 did not simply spill over into Eurasia, but was driven as part of an expansion that had begun in sub-Saharan Africa thousands of years earlier.

There are a number of possible explanations for such an expansion. A recent study of drill cores from Lake Malawi sediment indicate a change from highly variable arid conditions to a wetter, more stable climate in east Africa ~70kya (Scholz *et al.* 2007). The authors argue that such a climatic shift could have stimulated human population growth and later migrations from the region. While we find this argument persuasive, if climate was the only factor at work, the other haplogroups, also thought to be in east Africa at this time (Salas *et al.* 2002), should show a similar expansion pattern. We can explain the observed pattern of growth if an initially small group carrying the ancestral L3 haplotype gained a cultural advantage that allowed its population to out-compete rival groups. Consistent with this idea, Mellars (2006) has argued on the basis of archaeological evidence that the African exodus was pre-dated by a cultural revolution involving new stone blade technologies, skin working tools, ornaments and imported red ochre (although cf. Marean *et al.* [2007] for an earlier date). Red ochre,

also present in the earliest known human remains from Australia (Bowler & Thorne, 1976), is especially interesting in this context for its association with ritual and symbolism. More advanced symbolic systems in language and religious beliefs could have provided a competitive advantage to a group by promoting coordination and cohesion. Determining which, if any, of these factors were involved in the L3 expansion will require the synthesis of multiple lines of evidence from archaeology, paleoclimatology, historical linguistics and population genetics. Future work in population genetics, extending coalescent methods like those used here beyond a single locus to include Y-chromosome and genomic data, will further improve estimates of ancestral human population demographics and so help to provide an increasingly detailed understanding of the human story.

Acknowledgements

We acknowledge support from Leverhulme Trust grant F/00 239/T.

References

Adams, J. M. & Faure, H. 1997 Review and Atlas of Palaeovegetation: Preliminary land ecosystem maps of the world since the Last Glacial Maximum, vol. 2008 (ed. Q. members): Oak Ridge National Laboratory, TN, USA.

Alonso, S. & Armour, J. A. 2001 A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outstide Africa. *Proc. Natl. Acad. Sci. USA* **98**, 864-9.

Atkinson, Q. D., Gray, R. D. & Drummond, A. J. 2008 MtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. *Mol. Biol. Evol.* **25**, 468-474.

Bandelt, H. J., Alves-Silva, J., Guimaraes, P. E. M., Santos, M. S., Brehm, A., Pereira, L., Coppa, A., Larruga, J. M., Rengo, C., Scozzari, R., Torroni, A., Prata, M. J., Amorim, A., Prado, V. F. & Pena, S. D. J. 2001 Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Annals of Human Genetics* **65**, 549-563.

Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. 1995 Mitochondrial Portraits of Human-Populations Using Median Networks. *Genetics* **141**, 743-753.

Behar, D., 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, R., Rosset, S. & Consortium, G. 2008 The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1-11.

Bowler J. M. & Thorne, A. G. 1976 Human remains from Lake Mungo: Discovery and excavation of Lake Mungo III. In *The Origin of the Australians* (eds R. L. Kirk and A. G. Thorne), pp. 127–138. Canberra: Australian Institute of Aboriginal Studies.

Cann, R. L., Stoneking, M. & Wilson, A. C. 1987 Mitochondrial DNA and human evolution. *Nature* **325**, 31-6.

Chen, Y. S., Torroni, A., Excoffier, L., Santachiarabenerecetti, A. S. & Wallace, D. C. 1995 Analysis of Mtdna Variation in African Populations Reveals the Most Ancient of All Human Continent-Specific Haplogroups. *Am. J. of Hum. Genet.* **57**, 133-149.

Drummond, A. J. & Rambaut, A. 2003 BEAST. Version 1.3. http://evolve.zoc.ox.ac.uk/beast/. Oxford: Oxford University Press.

Drummond, A. J. & Rambaut, A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. 2005 Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185-92.

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.

Forster, P. & Matsumura, S. 2005 Evolution. Did early humans go north or south? *Science* **308**, 965-6.

Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G., Babb, P., Cerchio, S., Kaestle, F., Schanfield, M., Deka, R., Yanagihara, R. & Merriwether, D. A. 2005 Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Mol. Biol. Evol.* **22**, 1506-17.

Garrigan, D., Kingan, S. B., Pilkington, M. M., Wilder, J. A., Cox, M. P., Soodyall, H., Strassmann, B., Destro-Bisol, G., de Knijff, P., Novelletto, A., Friedlaender, J. & Hammer, M. F. 2007 Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics* **177**, 2195-2207.

Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. & Tishkoff, S. A. 2007 Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* **24**, 757-768.

Gonser, R., Donnelly, P., Nicholson, G. & Di Rienzo, A. 2000 Microsatellite mutations and inferences about human demography. *Genetics* **154**, 1793-807.

Harpending, H. & Rogers, A. 2000 Genetic perspectives on human origins and differentiation. *Annual Review of Genomics and Human Genetics* **1**, 361-385.

Harpending, H. C., Sherry, S. T., Rogers, A. R. & Stoneking, M. 1993 The Genetic-Structure of Ancient Human-Populations. *Current Anthropology* **34**, 483-496.

Hastings, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

Ho, S. Y., Phillips, M. J., Cooper, A. & Drummond, A. J. 2005 Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* **22**, 1561-8.

Ho, S. Y., *&* Endicott, P. 2008. The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am. J of Hum. Genet.* **83**, 142-146.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1-14.

Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. 2000 Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-13.

Kingman, J.F.C. 1982 On the Genealogy of Large Populations. *Journal of Applied Probability* **19A**, 27-43.

Kittler, R., Kayser, M. & Stoneking, M. 2003 Molecular evolution of Pediculds humanus and the origin of clothing. *Current Biology* **13**, 1414-1417.

Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P. A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L. L. & Oefner, P. J. 2006 The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373-87.

Knight, A., Underhill, P. A., Mortensen, H. M., Zhivotovsky, L. A., Lin, A. A., Henn, B. M., Louis, D., Ruhlen, M. & Mountain, J. L. 2003 African Y chromosome and mtDNA divergence provides insight into the history of click languages (vol 13, pg 464, 2003). *Current Biology* **13**, 705-705.

Maca-Meyer, N., Gonzalez, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. 2001 Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* **2**, 13.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N. K., Raja, J. M., 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**, 1034-6.

Marean, C. W., Bar-Matthews, M., Bernatchez, J., Fisher, E., Goldberg, P., Herries, A. I. R., Jacobs, Z., Jerardino, A., Karkanas, P., Minichillo, T., Nilssen, P. J., Thompson, E., Watts, I. & Williams, H. M. 2007 Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. *Nature* **449**, 905-U11.

Marth, G. T., Czabarka, E., Murvai, J. & Sherry, S. T. 2004 The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* **166**, 351-72.

Mellars, P. 2006 Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model (vol 103, pg 9381, 2006). *Proc. Natl. Acad. Sci. USA* **103**, 13560-13560.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087-1091.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., Sukernik, R. I., Olckers, A. & Wallace, D. C. 2003 Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* **100**, 171-6.

O'Connell, J. F. & Allen, J. 2004 Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. *Journal of Archaeological Science* **31**, 835-853.

Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D. M., Dugoujon, J. M., Coudray, C., Santachiara-Benerecetti, A. S., Semino, O., Bandelt, H. J. & Torroni, A. 2006 The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* **314**, 1767-70.

Rogers, A. R. 2001 Order emerging from chaos in human evolutionary genetics. *Proc. Natl. Acad. Sci. USA* **98**, 779-80.

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

Scholz, C. A., Johnson, T. C., Cohen, A. S., King, J. W., Peck, J. A., Overpeck, J. T., Talbot, M. R., Brown, E. T., Kalindekafe, L., Amoako, P. Y. O., Lyons, R. P., Shanahan, T. M., Castaneda, I. S., Heil, C. W., Forman, S. L., McHargue, L. R., Beuning, K. R., Gomez, J. & Pierson, J. 2007 East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proc. Natl. Acad. Sci. USA* **104**, 16416-16421.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., Kosintsev, P., Kunz, M. L., Martin, L. D., Stephenson, R. O., Storer, J., Tedford, R., Zimov, S. & Cooper, A. 2004 Rise and fall of the Beringian steppe bison. *Science* **306**, 1561-5.

Shapiro, B., Rambaut, A. & Drummond, A. J. 2006 Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23**, 7-9.

Sherry, S. T., Rogers, A. R., Harpending, H., Soodyall, H., Jenkins, T. & Stoneking, M. 1994 Mismatch distributions of mtDNA reveal recent human population expansions. *Hum Biol* **66**, 761-75.

Steel, M. A., Hendy, M. D. & Penny, D. 1988 Loss of information in genetic distances. *Nature* **336**, 118.

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A. G., Singh, V. K., Rasalkar, A. A. & Singh, L. 2005 Reconstructing the origin of Andaman Islanders. *Science* **308**, 996.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-82.

Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T. B., Ramakrishnan, U., Reed, F. A. & Mountain, J. L. 2007 History of click-speaking Populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.* **24**, 2180-2195.

Torroni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H. J. 2006 Harvesting the fruit of the human mtDNA tree. *Trends in Genetics* **22**, 339-345.

Underhill, P. A. & Kivisild, T. 2007 Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics* **41**, 539-564.

Underhill, P. A., Shen, P. D., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza,

A., Davis, R. W., Feldman, M. W., Cavalli-Sforza, L. L. & Oefner, P. J. 2000 Y chromosome sequence variation and the history of human populations. *Nature Genetics* **26**, 358-361.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. 1991 African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503-7.

Wakeley, J. 2000 The effects of subdivision on the genetic divergence of populations and species. *Evolution Int. J. Org. Evolution* **54**, 1092-101.

Watson, E., Forster, P., Richards, M. & Bandelt, H. J. 1997 Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* **61**, 691-704.

Witherspoon, D. J., Marchani, E. E., Watkins, W. S., Ostler, C. T., Wooding, S. P., Anders, B. A., Fowlkes, J. D., Boissinot, S., Furano, A. V., Ray, D. A., Rogers, A. R., Batzer, M. A. & Jorde, L. B. 2006 Human population genetic structure and diversity inferred from polymorphic L1 (LINE-1) and Alu insertions. *Human Heredity* **62**, 30-46.

Zhivotovsky, L. A., Bennett, L., Bowcock, A. M. & Feldman, M. W. 2000 Human population expansion and microsatellite variation. *Mol. Biol. Evol.* **17**, 757-67.

Figure Captions

Figure 1 – Genealogy of major African mtDNA haplogroups. This phylogeny shows the genealogical relationships between the L0, L1, L2 and L3 mtDNA lineages of Africa and the position of the two major non-African lineages, M and N.

Figure 2 – **BSPs of effective population size through time** for **a** Sub-Saharan Africa (n=224), **b** haplogroup L0 (n=60), **c** haplogroup L1 (n=41), **d** haplogroup L2 (n=43), and haplogroup L3 (n=80). The bold black line represents the median posterior effective population size through time. Finer lines delimit the 95% HPD for effective population size, accounting for uncertainty in the reconstructed phylogeny and substitution model parameters. The dotted line in 2a plots the sum of the separate lineage population estimates through time. Effective population size is plotted on a log scale and assumes a generation time of 20 years. These estimates of effective population size have an inverse relationship with the evolutionary rate of mtDNA used for the calibration, such that they will be lower for faster rates and higher for slower rates. For comparison, all the x-axis have a scale extending to 150kya. The plots are truncated to the median estimate of each region's TMRCA.

Figure 1

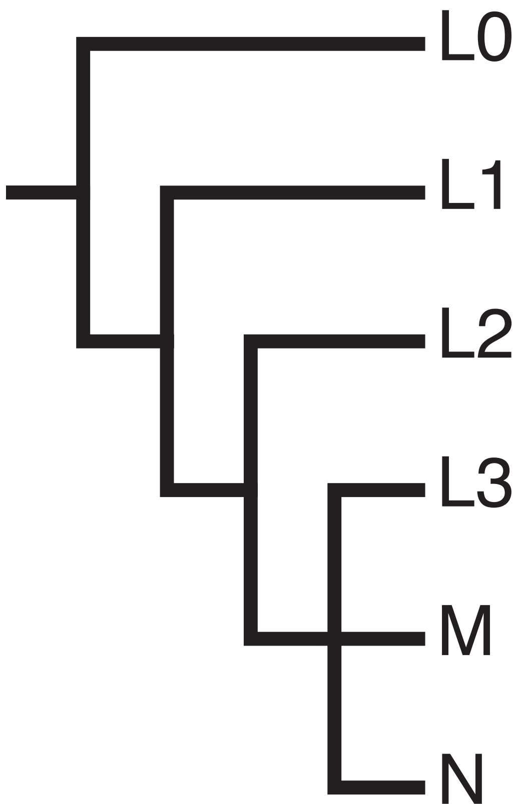

Figure 2