Atkinson, Q. D., Gray, R. D. and Drummond, A. J. (2008). mtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. Molecular Biology and Evolution, 25(2):468-474.

mtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory

Research Article

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

1 Institute of Cognitive and Evolutionary Anthropology, University of Oxford, 58A Banbury Rd, OX2 6QS, UNITED KINGDOM

2 Psychology Department, University of Auckland, Private Bag 92019, Auckland 1001, NEW ZEALAND

3 Bioinformatics Institute and Department of Computer Science, University of Auckland, Private Bag 92019, Auckland 1001, NEW ZEALAND

Name of corresponding author: - Quentin D. Atkinson

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

Telephone: - +44 (0)186 612 374 or +44 (0)789 482 8426

E-mail: - quentin.atkinson@anthro.ox.ac.uk

Keywords: Mitochondrial DNA, Human Evolution, Population genetics, Bayesian Skyline Plot, Anthropology.

Running Head: - mtDNA predicts human population size

The relative timing and size of regional human population growth following our expansion from Africa remains unknown. Human mitochondrial DNA (mtDNA) diversity carries a legacy of our population history. Given a set of sequences, we can use coalescent theory to estimate past population size through time and draw inferences about human population history. However, recent work has challenged the validity of using mtDNA diversity to infer species population sizes. Here we use Bayesian coalescent inference methods, together with a global dataset of 357 human mtDNA coding region sequences, to infer human population sizes through time across eight major geographic regions. Our estimates of relative population sizes show remarkable concordance with the contemporary regional distribution of humans across Africa, Eurasia and the Americas, indicating that mtDNA diversity is a good predictor of population size in humans. Plots of population size through time show slow growth in sub-Saharan Africa beginning 143-193kya, followed by a rapid expansion into Eurasia after the emergence of the first non-African mtDNA lineages 50-70kya. Outside Africa, the earliest and fastest growth is inferred in Southern Asia ~52kya, followed by a succession of growth phases in Northern and Central Asia (~49kya), Australia (~48kya), Europe (~42kya), the Middle East and North Africa (~40kya), New Guinea (~39kya), the Americas (~18kya), and a second expansion in Europe (~10-15kya). Comparisons of relative regional population sizes through time suggest that between approximately 45kya and 20kya most of humanity lived in Southern Asia. These findings not only support the use of mtDNA data for estimating human population size but also provide a unique picture of human prehistory and demonstrate the importance of Southern Asia to our recent evolutionary past.

Introduction

How humans colonized the globe remains 'one of the greatest untold stories in the history of humankind' (Goebel 2007). mtDNA has been a crucial line of evidence in developing the current understanding of our genetic prehistory. Phylogenetic studies of human mtDNA variation support a late Pleistocene expansion of modern humans from Africa (Cann, Stoneking, and Wilson 1987; Vigilant et al. 1991; Watson et al. 1997; Ingman et al. 2000). More recent work suggests a "southern" migration route from sub-Saharan Africa along the Indian Ocean coast into Eurasia (Forster and Matsumura 2005; Macaulay et al. 2005; Thangaraj et al. 2005) and Sahul (Ingman and Gyllensten 2003; Friedlaender et al. 2005; van Holst Pellekaan et al. 2006), and a later migration from the Levant into North Africa and Europe (Olivieri et al. 2006). However, despite these advances, we lack a clear understanding of the timing and size of regional human population expansions.

Past population size can be estimated from genetic data using coalescent theory (Hudson 1990). For any locus, the information that can be extracted about population size depends on the accuracy with which the underlying genealogical tree can be resolved, and the extent to which the historical demographic processes can be accommodated by the coalescent model. mtDNA's high copy number, absence of recombination (Elson et al. 2001), and rapid substitution rate (Howell, Kubacka, and Mackey 1996), mean that its genealogical tree is well-resolved, thus making mtDNA very well-suited for coalescent inference of population size. Recently, however, Bazin *et al.* (2006) have argued that mtDNA variation is a poor indicator of population size in animals, probably due to the effects of selection acting to reduce diversity (known

as genetic draft). Bazin *et al.* compared intra-species mtDNA and allozyme diversity across major taxonomic divisions and found no association between mtDNA diversity and population size, whereas allozyme diversity did predict population size. Whilst the pattern Bazin *et al.* (2006) identify appears not to hold for human genetic diversity (human mtDNA and autosomal diversity both imply a similar population size (Eyre-Walker 2006)) human mtDNA does show evidence of selection at some loci (Mishmar et al. 2003; Kivisild et al. 2006). This raises the question of whether mtDNA is in fact a reliable predictor of human population size.

Previous work has used coalescent analyses of pairwise mtDNA sequence differences to argue for exponential growth in human population size beginning 30-130kya (Harpending et al. 1993; Sherry et al. 1994; Rogers 1995; Ingman et al. 2000). Recent improvements in the available inference methods and complete mtDNA sequence data mean that population genetic parameters can now be estimated from sequence data with greater accuracy, without requiring a simple parametric growth model, and with an explicit framework for quantifying uncertainty in parameter estimates. The Bayesian Skyline Plot (BSP) (Drummond and Rambaut 2003; Drummond et al. 2005) uses bayesian coalescent inference of phylogeny and a Markov Chain Monte Carlo (MCMC) (Metropolis et al. 1953) sampling algorithm to simultaneously estimate a posterior probability distribution for the ancestral genealogy, branch lengths, substitution model parameters and population parameters through time, given a set of gene sequences. The resulting BSP represents a credibility interval for effective population size that incorporates uncertainty in nuisance parameters such as the underlying ancestral gene tree, branch lengths and rate parameters. Unlike previous coalescent inference methods, by generating a sample of piecewise population size

estimates, results are not contingent on a pre-specified parametric growth model and the posterior distribution can be used to directly assess uncertainty in parameter estimates given the assumptions of the model. In addition, information loss is minimized by inferring population parameters directly from the sequence data using a coalescent framework, rather than converting sequence data to pair-wise distances between sequences (Steel, Hendy, and Penny 1988).

In this report we use the Bayesian Skyline Plot together with a global dataset of 357 human mtDNA coding region sequences to estimate effective population sizes through time for eight major geographic regions. By comparing our coalescent-based estimates of regional effective population size with anthropological estimates of census population sizes we show that mtDNA diversity is a remarkably good predictor of regional human population size. As a result, we are able to use our regional population size estimates to gain novel insight into the human colonization of the globe during the last 100,000 years. Most strikingly, we find evidence of a major Southern Asian chapter in our recent evolutionary past.

Materials and Methods

Sequence Data. A dataset of 357 human mtDNA coding region sequences was assembled from the >3000 sequences available on Genbank and divided into 8 geographic regions (Table S1). Only sequences that could be unambiguously assigned to regional groupings were considered for analysis. We sought a balanced sample of between approximately 25 and 50 sequences from each of the regions, with a larger sampling of the more diverse African population. Data were machine aligned in ClustalX (Thompson et al. 1997) using reference sequence J01415.1. Regional

boundaries (figure 2a) were defined based on geographical barriers and independent genetic evidence for population subdivisions (Cavalli-Sforza, Menozzi, and Piazza 1994). Results were found to be robust across a range of different regional definitions (supplementary figures 1 - 3; Supplementary Material online). Alternative regional sequence groupings are shown in supplementary table 2.

Bayesian Skyline Plots. Bayesian Skyline Plots (BSP) (Drummond et al. 2005) of effective population size were produced from the data using MCMC (Metropolis et al. 1953) sampling in the programme *BEAST* (*version 1.3*) (Drummond and Rambaut 2003). Effective population size is a compound population genetic parameter generally considered linearly proportional to census population size – here, the population of breeding females. It is influenced by many factors, including local extinction and recolonization and various forms of non-random mating (Wakeley, 2000), and assumes that regional populations are isolated (i.e. there is no migration between populations). The underlying population size function of the BSP can be specified either as a stepwise or a piecewise linear function of population size change. The main figures were obtained with a piecewise linear model. Estimates of effective population size are derived from the inferred lineage coalescent rate through time. Ancestral gene trees were inferred based on a general time-reversible (GTR) substitution model with site-specific rates for 1st, 2nd and 3rd codon positions. A Bayes factor computed via importance sampling (Newton et al. 1994) indicated that this model fitted the data better than the commonly used GTR + Γ + I model (log Bayes factor = 19.1). 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 equilibrium distribution and provided adequate posterior samples by this time. All runs had an effective sample size of the coalescent prior of at least 1000.

In order to plot population size with respect to time it was necessary to calibrate rates of molecular evolution. Recent evidence of systematic variation in observed rates of molecular evolution through time (Ho et al. 2005) suggests that rates should be calibrated using divergence events of a similar time depth to the time scale of interest. For this reason an internal rate calibration was used. Rates were fixed to 1.691×10^{-8} substitutions/site/yr, based on a 45kya age for haplotype Q in New Guinea. Haplogroup Q occurs at high frequencies only in New Guinea (parts of Melanesia have a closely related subtype [Q2], which also occurs at very low frequencies in Micronesia and Polynesia, probably as a result of more recent admixture) [Friedlaender et al., 2005]). This is consistent with archaeological evidence for modern human arrival on the super continent Sahul (comprising what is now New Guinea and Australia) by approximately ~42-45kya (O'Connell and Allen 2004). In order to isolate uncertainty in the coalescent process and facilitate regional comparison, uncertainty in the calibration time itself was not incorporated into our analyses. Inferred times do, however, scale linearly with respect to the rate calibration, such that if, for example, a 50kya age for the New Guinea clade were assumed, the time scale would be increased by approximately 11%.

Results and Discussion

Figure 1 shows BSPs of effective population size through time for each of the eight regions examined here. The growth curves are derived from changes in the inferred

lineage coalescent rate through time. These changes are generally taken to reflect changes in census population size, but could also be influenced by changes in population structure, selection pressures or some combination of these. We can test whether our effective population size estimates provide a good indication of relative census population sizes by comparing relative size estimates for the 'present' with contemporary regional population sizes. Figure 2b shows that relative regional population size estimates at time t=0 are remarkably consistent with independent anthropological and historical estimates of relative pre-colonial population sizes across Africa, Eurasia and the Americas (Biraben 1979). Only estimates for Australia and New Guinea differ significantly from the anthropological estimates. This is likely to be due to differences in population level processes between Australia and New Guinea and the other six regions. Genetic distance between populations in Australia and New Guinea rises quickly with increasing geographic distance (Cavalli-Sforza, Menozzi, and Piazza 1994). Studies of Australian Aboriginals show that tribes are genetically very isolated (Huoponen et al. 2001), with the majority of marriages within the tribe (Birdsell 1993). Highly structured populations with low migration rates and little extinction/recolonization can be expected to preserve variation that would be lost due to drift in a larger meta-population and hence produce relatively high effective population size estimates (Wakeley 2000). Excluding Australia and New Guinea, there is a striking relationship between effective and census population size estimates ($r = 0.98$). This relationship was found to be robust across different definitions of the regional group boundaries (supplementary figures 1 and 2). These results indicate that mtDNA variation can be used to infer relative regional population sizes in humans.

The finding of Bazin et al. (2006) that population size does not influence species' mtDNA diversity across the animal kingdom appears not to apply, at least at the population level, in humans. This is most likely because the historically small human population size and short time scales involved make it unlikely that a selective sweep has been able to 'reset' regional mtDNA diversity since our expansion from Africa. We were therefore able to combine results from figure 1 to estimate the global population distribution of modern humans over the last 100,000 years (figure 2c). An important caveat in interpreting theses figures is that the population demographic history inferred from sequences sampled from a particular region does not necessarily represent the population history of the region, but rather the history of the genetic lineages now inhabiting the region. For example, regional interference can occur if a major migration event brings the genetic signature from a region with a long history to a newly inhabited region. However, we note that even in cases where we know this has happened, such as the expansion from Africa itself and the colonization of the Americas, our demographic estimates appear to correctly infer a recent expansion event rather than being mislead by the history of the parent population.

Figures 1 and 2 together provide a unique picture of the human colonization of the globe. The plots all show evidence of significant population growth and reveal clear differences in regional population histories. The time scale in these figures was calibrated by assuming a mtDNA evolutionary rate of 1.691×10^{-8} substitutions/site/yr. Using a slower rate would lead to older dates than presented. However, the relative times of the population expansions across regions would not be altered by a different calibration. We infer slow, roughly exponential population growth in Sub-Saharan Africa from an MRCA 143-193kya (95% HPD; Figure 1a),

followed by a period of rapid growth in Eurasia shortly after the emergence of the first non-African mtDNA lineages (N-haplogroup MRCA 53-69kya, 95% HPD; M haplogroup MRCA 50-64kya, 95% HPD). This result is consistent with analyses of human microsatellite and SNPs data suggesting slow growth in Africa and more recent rapid growth following a bottleneck in Asia and Europe (Gonser et al. 2000; Zhivotovsky et al. 2000; Alonso and Armour 2001; Rogers 2001; Marth et al. 2004). The congruence between our findings and those from unlinked genetic loci provides further evidence that the observed trends in effective population size are not merely an artefact of selection, which would require that identical selection pressures had acted simultaneously on unrelated genes.

Southern Asia shows the earliest and most pronounced population expansion outside Africa, with a five-fold increase in population size estimated by ~52kya (Figure 1b). An early Southern Asian growth phase is consistent with high mtDNA diversity in populations from the Indian Ocean Coast (Kivisild et al. 2003; Macaulay et al. 2005; Thangaraj et al. 2005; Sun et al. 2006), as well as recent archaeological evidence (Mellars 2006a), and strongly supports a rapid coastal migration along the "Southern Route" from Africa into Southern Asia. A southern expansion route is also supported by the analyses in Figure S1, with growth inferred first in South Western Asia (India/Middle East) and slightly later in South East Asia. Perhaps most striking, however, is the magnitude of the Southern Asian growth phase, which implies that between approximately 45kya and 20kya over half of the global human population lived on the Indian subcontinent and what is now the Thai and Malay Peninsulas. Our population size estimates indicate that the proportion of people living in the region peaked at over 60% approximately 38kya. This figure may even be an underestimate

given the potential for regional interference along the lines mentioned above, since early Southern Asian growth could cause us to infer a premature increase in the population size estimates for the other regions.

We infer less rapid, progressively later population growth in Northern and Central Asia (~49kya; Figure 1c), Europe (~42kya; Figure 1e) and the Middle East and North Africa (~40kya; Figure 1f). Slower growth rates in these regions may reflect less favourable conditions for human occupation at this time. The later timing of growth in Europe and the Middle East and North Africa is consistent with radio-carbon dating of the Upper Palaeolithic transition (Mellars 2006b) and mtDNA lineage analysis (Olivieri et al. 2006) indicating an expansion from the Levant into Europe and North Africa 40-45kya. We also see later growth in Australia (~48kya; Figure 1d) and New Guinea (~39kya; Figure 1g). In the Americas we find evidence of a rapid population expansion beginning ~18kya (Figure 1h). This result is consistent with a hypothesized colonization event after the last glacial maximum ~20kya, when lower sea levels made it possible to cross from Siberia to North America via the Bering land bridge (Forster 2004).

By ~35kya global growth outside Africa begins to slow, most markedly in Southern Asia and Australia, whilst in Europe estimated effective population size actually declines slightly. This levelling of growth may reflect the impact of Malthusian environmental constraints and, particularly in Europe, increasing glaciation during this time. Climate also seems the most likely explanation for a second major growth phase inferred in Europe beginning 10-15kya (Figure 1e), shortly after the end of the last ice age ~15kya. The estimated timing of this growth phase fits neatly with

archaeological evidence for an expansion of hunter-gatherers from glacial refugia in southern Europe (Housley et al. 1997). Such an interpretation is also supported by previous work into the relative Neolithic and Palaeolithic contributions to the European mtDNA gene pool (Richards et al. 2000) and the discovery amongst the Saami (in northern Scandinavia) of south-western European mtDNA haplotypes dated to the late Pleistocene (Achilli et al. 2005).

On a more recent time scale, the advent of agriculture is thought to have driven human population growth from ~5-10kya. Uncertainty in the timing of the second European expansion means we cannot rule out the possibility that this growth was linked, at least in part, to the spread of agriculture from Anatolia beginning ~9kya (Gkiasta et al. 2003; Gray and Atkinson 2003). However, the inferred expansion time is centred somewhat earlier than would be expected under this scenario and an equally strong growth signal is not inferred for Africa, Papua New Guinea, the Americas and the rest of Eurasia, where we would expect agriculture to have had a similar effect. This does not mean that no significant growth occurred with the development of early agriculture (although evidence of poor health in early agricultural populations makes this a possibility (Larsen 2006)). Rather, a more likely explanation is that due to the limits of the temporal resolution of the method, very recent expansion in human population size associated with agriculture has simply not left a growth signal that we can detect here. Whilst the resolution of the method is ultimately limited by the rate of molecular evolution, in the future multi-locus analyses will allow a finer-grained analysis.

The results reported here extend our understanding of the origin and evolution of modern humans. By verifying our regional population size estimates against independent anthropological and historical evidence we have shown that mtDNA variation can be used to predict population size in humans. Our estimates of effective population size through time show that Southern Asia was not only a key waypoint in the human expansion from Africa but a major chapter in human prehistory. In addition, we find complex patterns of regional population growth that can be linked with archaeological evidence and known climatic events. As the amount of available sequence data continues to grow and results from a range of loci are combined, we anticipate that coalescent-based methods will reveal an increasingly detailed picture of human prehistory.

Supplementary Material

Supplementary material is available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgements

We are grateful to Mark Beaumont, Athena Ferreira, Andrew Meade, Mark Pagel, Marcel van de Steeg and Chris Venditti for helpful advice and comments. We also thank the many contributors to Genbank.

Literature Cited

Achilli, A., C. Rengo, V. Battaglia, M. Pala, A. Olivieri, S. Fornarino, C. Magri, R. Scozzari, N. Babudri, A. S. Santachiara-Benerecetti, H. J. Bandelt, O. Semino,

and A. Torroni. 2005. Saami and Berbers--an unexpected mitochondrial DNA link. Am J Hum Genet **76**:883-886.

Alonso, S., and J. A. Armour. 2001. A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outstide Africa. Proc Natl Acad Sci U S A **98**:864-869.

Bazin, E., S. Glemin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science **312**:570-572.

Biraben, J. N. 1979. Essay on the Evolution of Numbers of Mankind. Population **34**:13-25.

Birdsell, J. B. 1993. Microevolutionary patterns in Aboriginal Australia : a gradient analysis of clines. Oxford University Press, New York.

Cann, R. L., M. Stoneking, and A. C. Wilson. 1987. Mitochondrial DNA and human evolution. Nature **325**:31-36.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. The History and Geography of Human Genes. Princeton University Press, Princeton, NJ.

Drummond, A. J., and A. Rambaut. 2003. BEAST. Version 1.3. available at http://beast.bio.ed.ac.uk/

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol **22**:1185-1192.

Elson, J. L., R. M. Andrews, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, and N. Howell. 2001. Analysis of European mtDNAs for recombination. American Journal of Human Genetics **68**:145-153.

Eyre-Walker, A. 2006. Evolution - Size does not matter for mitochondrial DNA. Science **312**:537-538.

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; discussion 264.

Forster, P., and S. Matsumura. 2005. Evolution. Did early humans go north or south? Science **308**:965-966.

Friedlaender, J., T. Schurr, F. Gentz, G. Koki, F. Friedlaender, G. Horvat, P. Babb, S. Cerchio, F. Kaestle, M. Schanfield, R. Deka, R. Yanagihara, and D. A. Merriwether. 2005. Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol Biol Evol **22**:1506-1517.

Gkiasta, M., T. Russell, S. Shennan, and J. Steele. 2003. Neolithic transition in Europe: The radiocarbon record revisited. Antiquity **77**:45-62.

Goebel, T. 2007. Anthropology. The missing years for modern humans. Science **315**:194-196.

Gonser, R., P. Donnelly, G. Nicholson, and A. Di Rienzo. 2000. Microsatellite mutations and inferences about human demography. Genetics **154**:1793-1807.

Gray, R. D., and Q. D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature **426**:435-439.

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

Ho, S. Y., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol **22**:1561-1568.

Housley, R. A., C. S. Gamble, M. Street, and P. Pettitt. 1997. Radiocarbon evidence for the Lateglacial human recolonisation of northern Europe. Proceedings of the Prehistoric Society **63**:25-26.

Howell, N., I. Kubacka, and D. A. Mackey. 1996. How rapidly does the human mitochondrial genome evolve? American Journal of Human Genetics **59**:501-509.

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

Huoponen, K., T. G. Schurr, Y. Chen, and D. C. Wallace. 2001. Mitochondrial DNA variation in an aboriginal Australian population: evidence for genetic isolation and regional differentiation. Hum Immunol **62**:954-969.

Ingman, M., and U. Gyllensten. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res **13**:1600-1606.

Ingman, M., H. Kaessmann, S. Paabo, and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. Nature **408**:708-713.

Kivisild, T., S. Rootsi, M. Metspalu, S. Mastana, K. Kaldma, J. Parik, E. Metspalu, M. Adojaan, H. V. Tolk, V. Stepanov, M. Golge, E. Usanga, S. S. Papiha, C. Cinnioglu, R. King, L. Cavalli-Sforza, P. A. Underhill, and R. Villems. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet **72**:313-332.

Kivisild, T., P. Shen, D. P. Wall, B. Do, R. Sung, K. Davis, G. Passarino, P. A. Underhill, C. Scharfe, A. Torroni, R. Scozzari, D. Modiano, A. Coppa, P. de Knijff, M. Feldman, L. L. Cavalli-Sforza, and P. J. Oefner. 2006. The role of selection in the evolution of human mitochondrial genomes. Genetics **172**:373-387.

Larsen, C. S. 2006. The agricultural revolution as environmental catastrophe: Implications for health and lifestyle in the Holocene. Quaternary International **150**:12-20.

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

Marth, G. T., E. Czabarka, J. Murvai, and S. T. Sherry. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics **166**:351-372.

Mellars, P. 2006a. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science **313**:796-800.

Mellars, P. 2006b. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. Nature **439**:931-935.

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

Mishmar, D., E. Ruiz-Pesini, P. Golik, V. Macaulay, A. G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, M. D. Brown, R. I. Sukernik, A. Olckers, and D. C. Wallace. 2003. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A **100**:171-176.

Newton, M. A., A. E. Raftery, A. C. Davison, M. Bacha, G. Celeux, B. P. Carlin, P. Clifford, C. Lu, M. Sherman, M. A. Tanner, A. E. Gelfand, B. K. Mallick, A. Gelman, A. P. Grieve, H. R. Kunsch, T. Leonard, J. S. J. Hsu, J. S. Liu, D. B. Rubin, A. Y. Lo, T. A. Louis, R. M. Neal, A. B. Owen, D. S. Tu, W. R. Gilks, G. Roberts, T. Sweeting, D. Bates, G. Ritter, B. J. Worton, G. A. Barnard, R. Gibbens, and B. Silverman. 1994. Approximate Bayesian-Inference with the Weighted Likelihood Bootstrap. Journal of the Royal Statistical Society Series B-Methodological **56**:3-48.

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

Olivieri, A., A. Achilli, M. Pala, V. Battaglia, S. Fornarino, N. Al-Zahery, R. Scozzari, F. Cruciani, D. M. Behar, J. M. Dugoujon, C. Coudray, A. S. Santachiara-Benerecetti, O. Semino, H. J. Bandelt, and A. Torroni. 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science **314**:1767-1770.

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

Rogers, A. R. 1995. Genetic-Evidence for a Pleistocene Population Explosion. Evolution **49**:608-615.

Rogers, A. R. 2001. Order emerging from chaos in human evolutionary genetics. Proc Natl Acad Sci U S A **98**:779-780.

Sherry, S. T., A. R. Rogers, H. Harpending, H. Soodyall, T. Jenkins, and M. Stoneking. 1994. Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol **66**:761-775.

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

Sun, C., Q. P. Kong, M. G. Palanichamy, S. Agrawal, H. J. Bandelt, Y. G. Yao, F. Khan, C. L. Zhu, T. K. Chaudhuri, and Y. P. Zhang. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol **23**:683-690.

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

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25**:4876-4882.

van Holst Pellekaan, S. M., M. Ingman, J. Roberts-Thomson, and R. M. Harding. 2006. Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. Am J Phys Anthropol **131**:282-294.

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

Wakeley, J. 2000. The effects of subdivision on the genetic divergence of populations and species. Evolution Int J Org Evolution **54**:1092-1101.

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

Zhivotovsky, L. A., L. Bennett, A. M. Bowcock, and M. W. Feldman. 2000. Human population expansion and microsatellite variation. Mol Biol Evol **17**:757-767.

Figure 1 – Bayesian Skyline Plots of effective population size through time

a Sub-Saharan Africa, **b** South Asia, **c** Northern and Central Asia, **d** Australia, **e** Europe, **f** the Middle East and North Africa, **g** New Guinea, and **h** the Americas. The black line represents the median posterior effective population size through time. Blue lines bound the 95% Highest Posterior Density (HPD) for effective population size, accounting for uncertainty in the reconstructed phylogeny and substitution model parameters. 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 80kya (except for Africa which extends to 160kya). The plots are truncated to the median estimate of each region's most recent common ancestor (MRCA). Due to uncertainty in the time to coalescence, these MRCA times are not in themselves a good indicator of the colonization time of each region.

Figure 2 – Comparative analysis of relative regional population sizes through time.

a Color coded map showing the eight geographic regions – Sub-Saharan Africa (maroon), Southern Asia (dark blue), Northern and Central Asia (turquoise), Europe (green), the Middle East and North Africa (yellow), the Americas (orange), and Australia and New Guinea (pink). Grey regions were not sampled. **b** Comparison of anthropological estimates (Biraben 1979) of relative regional population size (hatched bars) and effective population size estimates at t=0 (solid bars), color coded as in panel **a**. Anthropological estimates are taken from Biraben (1979) combine population

density estimates with early census data. Estimates at 1600AD were used rather than modern estimates to avoid the effects of colonialism, which greatly altered relative regional population sizes. Anthropological and genetic estimates are scaled relative to the combined total population excluding Australia and New Guinea. **c** Cumulative plot of human population size through time for six geographic regions, color coded as in panel **a**. Australia and New Guinea were not included in the cumulative plot as effective population size estimates for these regions were deemed unreliable.

Figure 1 -

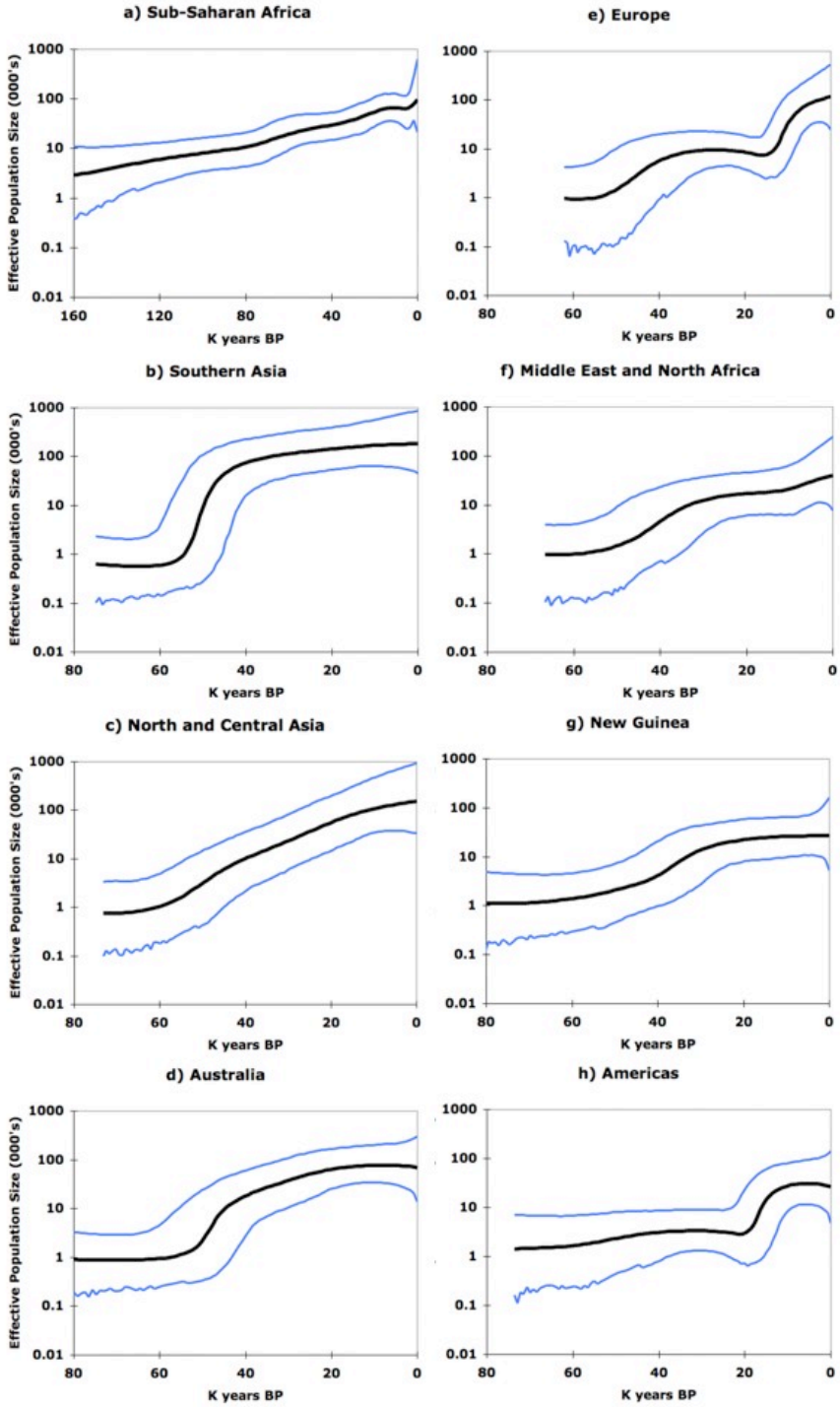
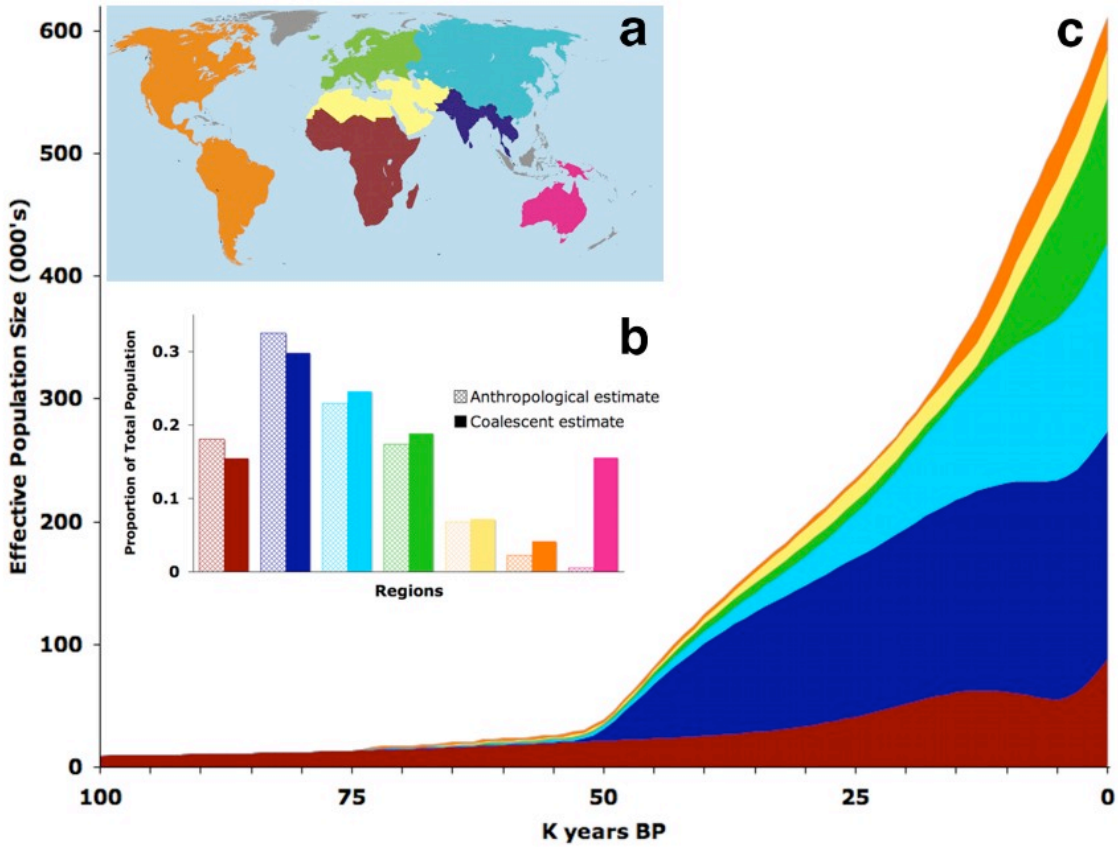

a) Sub-Saharan Africa
e) Europe
b) Southern Asia
f) Middle East and North Africa
c) North and Central Asia
g) New Guinea
d) Australia
h) Americas
Effective Population Size (000's)
1000
100
10
1
0.1
0.01
160
120
80
40
60
20
0
K years BP

Figure 2 –


a
b
c
Effective Population Size (000's)
600
500
400
300
200
100
0
100
75
50
25
0
K years BP
Proportion of Total Population
0.3
0.2
0.1
0
Regions
Anthropological estimate
Coalescent estimate