Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages

Mary Katherine Gonder, Holly M. Mortensen,* Floyd A. Reed,* Alexandra de Sousa,†‡ and Sarah A. Tishkoff**

*Department of Biology, University of Maryland; †Center for the Advanced Study of Hominid Paleobiology and ‡Department of Anthropology, George Washington University

Studies of human mitochondrial (mt) DNA genomes demonstrate that the root of the human phylogenetic tree occurs in Africa. Although 2 mtDNA lineages with an African origin (haplogroups M and N) were the progenitors of all non-African haplogroups, macrohaplogroup L (including haplogroups L0–L6) is limited to sub-Saharan Africa. Several L haplogroup lineages occur most frequently in eastern Africa (e.g., L0a, L0f, L5, and L3g), but some are specific to certain ethnic groups, such as haplogroup lineages L0d and L0k that previously have been found nearly exclusively among southern African "click" speakers. Few studies have included multiple mtDNA genome samples belonging to haplogroups that occur in eastern and southern Africa but are rare or absent elsewhere. This lack of sampling in eastern Africa makes it difficult to infer relationships among mtDNA haplogroups or to examine events that occurred early in human history. We sequenced 62 complete mtDNA genomes of ethnically diverse Tanzanians, southern African Khoisan speakers, and Bakola Pygmies and compared them with a global pool of 226 mtDNA genomes. From these, we infer phylogenetic relationships amongst mtDNA haplogroups and estimate the time to most recent common ancestor (TMRCA) for haplogroup lineages. These data suggest that Tanzanians have high genetic diversity and possess ancient mtDNA haplogroups, some of which are either rare (L0d and L5) or absent (L0f) in other regions of Africa. We propose that a large and diverse human population has persisted in eastern Africa and that eastern Africa may have been an ancient source of dispersion of modern humans both within and outside of Africa.

Introduction

Genetic analysis of mitochondrial (mt) DNA has been an important tool in understanding human evolution due to characteristics of mtDNA, such as high copy number, lack of recombination, high substitution rate, and a maternal mode of inheritance (Ballard and Whitlock 2004). However, most studies of human evolution that have included mtDNA sequences have been confined to the d-loop, which occupies less than 7% of the mtDNA genome (e.g., Cann et al. 1987; Vigilant et al. 1991; Chen et al. 1995; Watson et al. 1996; Quintana-Murci et al. 1999; Wallace et al. 1999; Chen et al. 2000; Jorde et al. 2000; Salas et al. 2002; Kivisild et al. 2004; Salas et al. 2004). Inferences drawn from the d-loop alone can be problematic given that the d-loop mutates rapidly and is subject to saturation due to excessive homoplasy (Tamura and Nei 1993; Bandelt et al. 2006). Mutations are not randomly distributed across the length of the locus, making rate heterogeneity another important issue in calculating divergence date estimates (Excoffier and Yang 1999; Meyer et al. 1999; Bandelt et al. 2006). In addition, several equally likely gene trees can often be inferred from d-loop sequences, particularly when large numbers of samples are analyzed (Maddison et al. 1992; Ballard and Whitlock 2004).

Comprehensive studies of the human mtDNA genome have been carried out by analyzing single nucleotide polymorphisms (SNPs) determined by restriction fragment length polymorphism (RFLP) analysis and sequences of the first hypervariable region of the d-loop (Chen et al. 1995, 2000; Salas et al. 2002, 2004). These studies have demonstrated that human mtDNA is geographically structured and may be classified into groups of related haplotypes (i.e, haplogroups) (Chen et al. 1995; Wallace et al. 1999). Only 2 mtDNA macrohaplogroups (M and N) and their derivatives persisted in non-Africans after the migration of modern humans out of Africa. Macrohaplogroup L is geographically limited to sub-Saharan Africa and has been divided into haplogroups L0–L6 (Mishmar et al. 2003; Salas et al. 2004; Kivisild, Metspalu, et al. 2006). The phylogeny of macrohaplogroup L is largely based on d-loop sequence and RFLP analysis and is, therefore, not well resolved (fig. 1*A* and *B*) particularly at basal tree nodes (Kivisild, Metspalu, et al. 2006). In particular, African mtDNAs that belong to L0 and L1 fall into several distinctive subhaplogroups, but their history is complex and poorly understood (Pereira et al. 2001; Kivisild, Metspalu, et al. 2006).

Haplogroup L0 is divided into subhaplogroups L0a, L0d, L0f, and L0k (Salas et al. 2002; Mishmar et al. 2003; Salas et al. 2004; Kivisild, Metspalu, et al. 2006) (fig. 1). Based on their geographic distribution throughout Africa (fig. 1*C*), Salas et al. (2002, 2004) suggested the following scenario about the origin and diversification of subhaplogroups belonging to L0 and L1. L0a probably originated in eastern Africa and is common in eastern, central, and southeastern Africa, but is almost absent in northern, western, and southern Africa (Salas et al. 2002). L0d is found almost exclusively among southern African Khoisan (SAK) speakers (Salas et al. 2002). There is only a single L0d haplotype from a Turkana of Kenya (Watson et al. 1997), as well as a few L0d haplotypes from Mozambique, which may have arisen by recent gene flow with the SAK (Pereira et al. 2001). Similarly, L0k is found exclusively among the SAK (Salas et al. 2002). However, the phylogenetic relationship of L0d and L0k is uncertain (Watson et al. 1997; Kivisild, Metspalu, et al. 2006). L0f is rare and appears to be geographically confined to east Africa (Salas et al. 2002). Haplogroup L1 is composed of L1b and L1c (fig. 1*C*). L1b is concentrated in western Africa, but it also occurs in central and northern Africa (Watson et al. 1997; Rosa et al. 2004). L1c occurs frequently among central African Bantu speakers (Vigilant et al. 1991; Destro-Bisol et al. 2004) and probably originated among peoples near the Atlantic coast in western equatorial

Key words: genetics, mtDNA genomes, Africa, mtDNA haplogroups, *Homo sapiens*, Tanzania, Khoisan speakers.

E-mail: tishkoff@umd.edu.

Mol. Biol. Evol. 24(3):757–768. 2007
doi:10.1093/molbev/msl209
Advance Access publication December 28, 2006

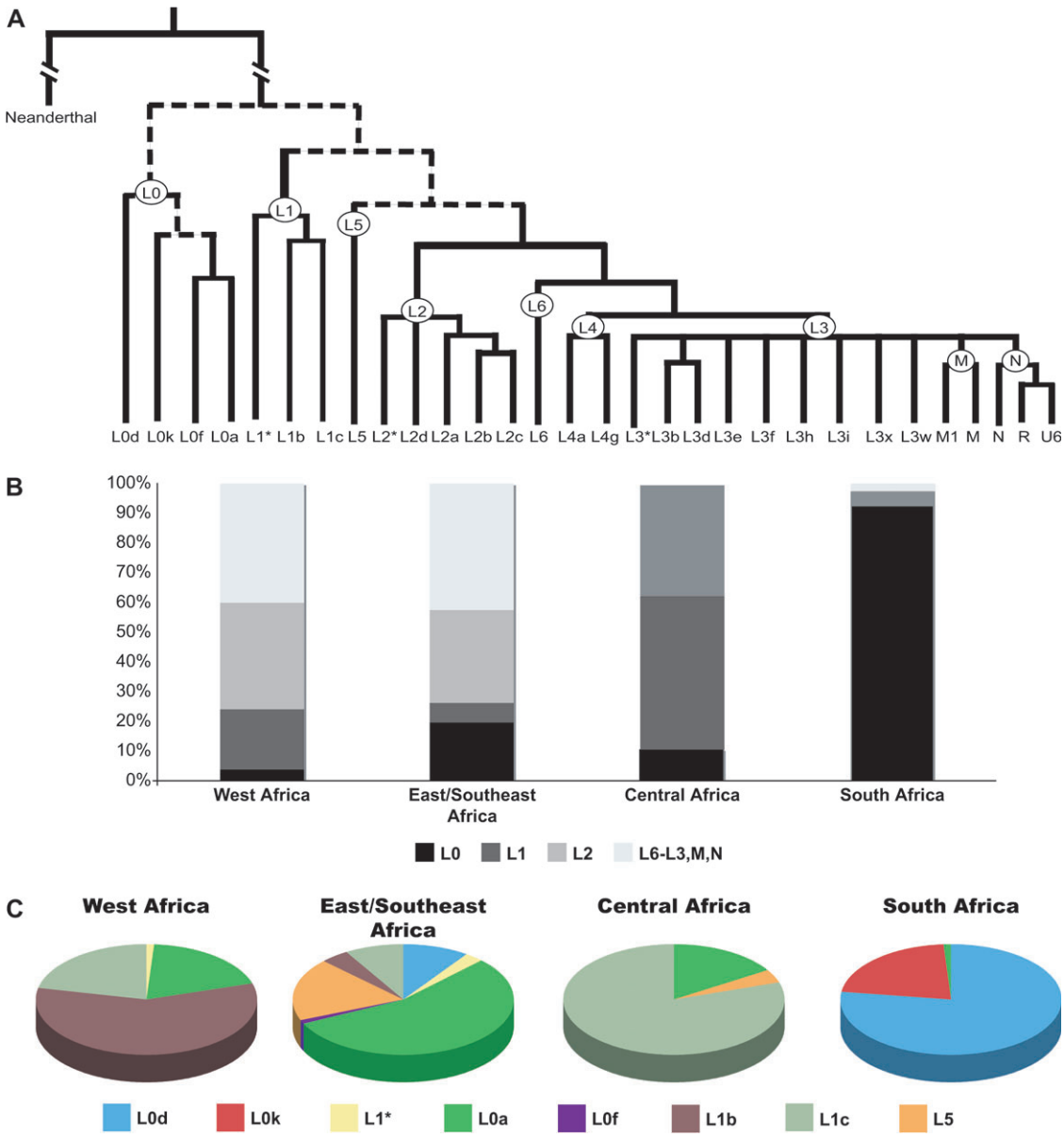

FIG. 1.—Evolutionary history of mtDNA haplogroup structure in African populations inferred from mtDNA d-loop and RFLP analysis. (*A*) Relationships among different mtDNA haplogroup lineages inferred from mtDNA d-loop sequences and mtDNA coding region SNPs from previous studies (Kivisild, Metspalu, et al. 2006). Dashed lines indicate previously unresolved relationships. (*B*) Relative frequencies of haplogroups L0, L1, L5, L2, L3, M, and N in different regions of Africa from mtDNA d-loop and mtDNA coding region SNPs from previous studies. (*C*) Relative frequencies of haplogroups L0, L1, and L5 subhaplogroups (excluding L2 and L3) in different regions of Africa from mtDNA d-loop and mtDNA coding region SNPs from previous studies. Haplogroup frequencies from previously published studies include East Africans (Ethiopia [Rosa et al. 2004], Kenya and Sudan [Watson et al. 1997; Rosa et al. 2004]), Mozambique (Pereira et al. 2001; Salas et al. 2002), Hadza (Vigilant et al. 1991), and Sukuma (Knight et al. 2003); South Africans (Botswana !Kung [Vigilant et al. 1991]); Central Africans (Mbenzele Pygmies [Destro-Bisol et al. 2004], Biaka Pygmies [Vigilant et al. 1991], and Mbuti Pygmies [Vigilant et al. 1991]); West Africans (Niger, Nigeria [Vigilant et al. 1991; Watson et al. 1997]; and Guinea [Rosa et al. 2004]). L1*, L2*, and L3* from previous studies indicate samples that were not further subdivided into subhaplogroups.

Africa (Salas et al. 2002, 2004). Both L1b and L1c are nearly absent in eastern and southern Africa. Haplogroup L5 (Kivisild et al. 2004) (previously referred to as L1e [Pereira et al. 2001]) has been observed at low frequency only in eastern Africa (Salas et al. 2002), Egypt (Stevanovitch et al. 2004), and among the Mbuti Pygmies (Kivisild, Shen, et al. 2006).

A global sample of complete mtDNA genome sequences has become publicly available making it possible to more precisely make phylogenetic inferences and calculate divergence dates for these mtDNA haplogroups (Ingman et al. 2000; Torroni et al. 2001; Ingman and Gyllensten 2003; Mishmar et al. 2004; Ruiz-Pesini et al. 2004; Macaulay et al. 2005; Thangaraj et al. 2005; Kivisild, Metspalu, et al. 2006; Kivisild, Shen, et al. 2006). However, these previous analyses have included few samples representing many of the African mtDNA subhaplogroups (L0a, L0d, L0f, L0k, L1b, L1c, and L5), particularly from people residing in eastern Africa. East African populations may provide important clues toward understanding modern

human origins. Both paleobiological and archeological data indicate that modern humans may have originated in eastern Africa (McBrearty and Brooks 2000; White et al. 2003), perhaps as early as 196,000 years ago (kya) (McDougall et al. 2005). In addition, the earliest migrations of modern humans out of Africa are thought to have originated from eastern Africa (Tishkoff et al. 1996; Quintana-Murci et al. 1999; Kivisild et al. 2004). Despite the paleobiological evidence that modern humans originated in eastern Africa, previous genetic studies have observed that L0k and L0d, which are found primarily among the SAK, occur at the root of the human mtDNA gene tree (Chen et al. 1995; Ingman et al. 2000; Mishmar et al. 2003; Ruiz-Pesini et al. 2004; Kivisild, Metspalu, et al. 2006). However, the presence of click-speaking populations in Tanzania (the Hadza and Sandawe) as well as Y chromosome data from the Hadza (Knight et al. 2003), Ethiopian, and Sudanese populations (Underhill et al. 2000, 2001; Cruciani et al. 2002; Semino et al. 2002) indicate that the SAK may have originated in eastern Africa, although the divergence between populations from these regions was quite ancient. Until now, no genetic data existed for the Sandawe.

In this study, we compare several complete mtDNA genomes of Tanzanians with a global panel of mtDNA genomes to clarify the evolutionary history of the African mtDNA haplogroups and to better characterize the role that populations in East Africa played in the origin and dispersal of modern humans across Africa. Generally, Tanzanians appear to have a high level of mtDNA genome diversity that is distributed among several mtDNA haplogroups that originated at different times in modern human history. These data suggest that populations in Tanzania have played an important and persistent role in the origin and diversification of modern humans.

Materials and Methods

Sample Collection

All Tanzanian samples were obtained from blood samples collected with informed consent and Institutional Review Board approval. The Ju-speaking !Xun (also known as Vasekela) and Khoe-speaking Khwe samples were collected from individuals in the area of Schmidtsdrift in the northwest Cape of South Africa and were provided by Dr M. Kotze. We obtained additional SAK samples from the Human Genome Diversity Panel–Centre d'Etude du Polymorphisme Humain (HGDP–CEPH). We selected samples for this study in 3 ways. First, we chose a sample of Tanzanians from 5 linguistically and culturally diverse ethnic groups (language classification is listed within parentheses): Sandawe (Khoisan), Hadza (Khoisan), Burunge (Afro-Asiatic), Maasai (Nilo-Saharan), and Turu (Niger-Kordofanian). Samples were chosen to represent the relative frequencies of the mtDNA L2, L3, M, and N haplogroups present in a much larger sample of d-loop sequences and mtDNA SNP data collected from over 700 Tanzanians (Gonder MK, Mortensen H, Reed F, Tishkoff SA, unpublished data). Second, we sequenced a subset of 25 Tanzanian samples to represent the most ancient mtDNA L0, L1, and L5 haplogroups (L0a, L0d, L0f, L1c, and L5). Third, we sequenced 10 samples from the SAK, 9 of which belong to mtDNA haplogroups L0d ($n = 7$) and L0k ($n = 2$). Finally, we sequenced the mtDNA genomes of 4 Bakola pygmies to expand our sample size of haplogroup L1c. All samples were combined for analysis with a global data set of 254 human sequences obtained from mtDB—Human Mitochondrial Genome Database (http://www.genpat.uu.se/mtDB/) and from GenBank. Complete mtDNA genome sequences of a chimpanzee (*Pan troglodytes*) and a gorilla were used as outgroups to the human mtDNA genomes for phylogenetic analyses (GenBank accession numbers D38113 and X93347).

Sequencing

We amplified mtDNA genome sequences in two 8.5 kilobase (kb) using overlapping fragments, a touchdown polymerase chain reaction (PCR) protocol (Don et al. 1991) and high-fidelity Platinum *Taq* polymerase following the manufacturer's protocol (Invitrogen Corporation, Carlsbad, CA). Sequencing was done using Big Dye Ready Reactions Kits using protocols specified by the manufacturer (Applied Biosystems, Inc., Foster City, CA). We processed 48 sequencing reactions using an ABI 3100 Genetic Analyzer for each individual, resulting in complete upstream and downstream mtDNA genome sequences (Rieder et al. 1998). PCR and sequencing primer sequences are given in table S1 (Supplementary Material online). We assembled sequences using Sequencher 4.1 (GeneCodes Corporation, Ann Arbor, MI) and annotated them according to the Cambridge Reference Sequence (Andrews et al. 1999). We prepared the mtDNA sequences for phylogenetic analysis in ClustalX (Thompson et al. 1997) by aligning them to the mtDNA genome sequences from mtDB Web site. We improved the resulting alignment by visual inspection in MacClade version 4.05 (Maddison WP and Maddison DR 2000).

Statistical Tests of Genetic Diversity and Neutrality

Because the d-loop is prone to problems with homoplasy and has been shown to produce unreliable gene trees (Maddison et al. 1992; Ballard and Whitlock 2004), we excluded the d-loop from all analyses shown in this study (except fig. S1, Supplementary Material online). We included sites corresponding to basepairs 577–16,023 of the Cambridge Reference Sequence (Andrews et al. 1999) in our analysis. We calculated the following summary statistics using DnaSP version 3.99 (Rozas et al. 2003): numbers of sequences (n), segregating sites (S), nucleotide diversity (π), and average number of nucleotide differences (k) for various subsets of the mtDNA genome sequences (Rozas et al. 2003). We also tested for deviations from expectations of neutrality, including Tajima's D, D^* of Fu and Li, and F^* of Fu and Li using DnaSP.

Phylogenetic Analyses

We determined optimal models of nucleotide sequence evolution by log likelihood ratio tests (Huelsenbeck and Crandall 1997) as implemented by PAUP* version 4.0b10 (Swofford 2002). We calculated a distance matrix and Neighbor-Joining (NJ) tree for the 324 samples using the HKY85 substitution model, with gamma-distributed rates

Table 1
Summary Statistics and Neutrality Tests of mtDNA Genomes (excluding the d-loop)

	n[a]	S[b]	$\pi \times 10^{-3}$[c]	k[d]	Tajima's D	D^* of Fu and Li	F^* of Fu and Li
Global	320	1,545	2.69	41.4	−2.60*	−7.74**	−5.82**
African	94	758	3.92	60.3	−2.05*	−3.97**	−3.75**
Non-African	226	1,086	1.81	27.9	−2.71***	−7.81**	−6.15**
Tanzanian	49	483	3.80	58.5	−1.70	−2.55*	−2.66*

[a] Numbers of sequences.
[b] Segregating sites.
[c] Nucleotide diversity.
[d] Average number of nucleotide differences (k) for mtDNA genomes that correspond to basepairs 577–16,023 of the Cambridge Reference Sequence.
*$P < 0.05$; **$P < 0.02$; ***$P < 0.001$.

and 32 discrete nucleotide substitution rate categories. We subjected the resulting tree to 100,000 bootstrap replicates with resampling to provide statistical support for the basal branches of the mtDNA gene tree.

In order to better resolve the phylogenetic relationships of the African mtDNA lineages, we also analyzed a subset of the original data set using Bayesian analysis. Due to computational limitations, this data set included all mtDNA genome sequences from sub-Saharan Africans ($n = 89$), African representatives of the M and N haplogroups ($n = 5$), and a geographically diverse sample of mtDNAs of non-Africans derived from haplogroups M and N ($n = 18$). We completed phylogenetic analyses of this smaller data set in MrBayes 3.1 (Ronquist and Huelsenbeck 2003) using the optimal model of sequence evolution determined in PAUP* (i.e., HKY85). In MrBayes, we ran 4 chains simultaneously for one million generations until the standard deviation (SD) of split frequencies was less than 0.01 under the HKY85 model with gamma-distributed rates and 16 rate categories.

Time to Most Recent Common Ancestor

We tested for nucleotide substitution rate heterogeneity for each gene tree using a likelihood ratio test ($-2\ln\Lambda$) comparing constrained (molecular clock enforced) versus unconstrained (no clock) trees (Huelsenbeck and Crandall 1997). Because we found significant rate heterogeneity in the NJ and Bayesian trees, we estimated divergence times for the mtDNA haplogroup clades using a penalized likelihood (PL) model as implemented in the program r8s 1.07 (Sanderson 1997, 2002, 2003) using the optimal smoothing value ($S = 320$) obtained by a cross-validation procedure in R8s.

We estimated confidence intervals (CIs) for each tree node using a 100 replicate bootstrap resampling procedure (Baldwin and Sanderson 1998) that was implemented by Perl scripts in the r8s-bootkit provided by Torsten Eriksson at http://www.bergianska.se/index_forskning_soft.html. We generated 100 bootstrap replicate data sets from the tree obtained from MrBayes v. 3.1 using SEQBOOT in PHYLIP (Felenstein 1993). In order to determine the 95% CI, we used a log likelihood decline of 2.0 units, which is roughly equivalent to 2 SDs (Sanderson and Doyle 2001). We conducted all analyses using a discrete approximation of a gamma distribution to accommodate for among-site rate heterogeneity. We calibrated our time to most recent common ancestor (TMRCA) estimates by assuming that the *Pan* and *Homo* lineages had separated from each other completely by 6 MYA (Kumar et al. 2005; Patterson et al. 2006) and added 500 ky for lineage sorting (Macaulay et al. 2005).

Median-Joining Network Analysis

Networks of L0/L1 mtDNA genome haplogroups were constructed using Network 4.1.1.1 (Fluxus Technology Ltd., 2004 [Bandelt et al. 1999]) in order to provide a detailed analysis of nucleotide substitutions along branches. We prepared sequences for analysis in MacClade 4.06 OS X. We excluded all invariant nucleotide positions in our L0/L1 alignment. We found 486 variable sites in the coding region, which spanned basepairs 593–16,077 of the Cambridge Reference Sequence (Andrews et al. 1999). These 486 variable sites were each assigned equal weight in the analyses (for contrasting method see Finnila et al. 2001). Additionally, we constructed a median-joining (MJ) network including *P. troglodytes* as an outgroup (GenBank accession number D38113). Although Network 4.1.1.1 is not intended for interspecies comparison, the chimpanzee was included to root the network and to unambiguously infer branching patterns at the base of the human mtDNA network for comparisons to our phylogenetic analyses.

Results

We sequenced a total of 62 African complete mtDNA genomes for this study that have been assigned GenBank accession numbers EF184580–EF184641. These mtDNA genomes were from individuals belonging to several ethnic groups in Tanzania ($n = 49$), click-speaking !Xun and Khwe populations from South Africa ($n = 10$), and Bakola Pygmies from Cameroon ($n = 4$). These samples were selected in order to fully represent L0a, L0d, L0f, L0k, L1c, L5, L2, L3, M, and N haplogroup lineages (table S2, Supplementary Material online). These newly sequenced African mtDNA genomes were aligned and compared with a global assortment of 254 mtDNA genomes of peoples of diverse geographic origin. Table S3 (Supplementary Material online) lists the GenBank accession number, sampling provenance, and major geographic region for these 254 mtDNAs. Diversity statistics are given in table 1. The genetic diversity present in this sample was broadly consistent with previous studies (Ingman et al. 2000; Ruiz-Pesini et al. 2004). The amount of mtDNA sequence diversity (π) among Africans (3.92×10^{-3}) and Tanzanians (3.80×10^{-3}) was more than twice that among non-Africans (1.81×10^{-3}). However, the level of variation in Africa may be artificially

elevated to some extent by the selection of genomes for sequencing that would maximize haplogroup representation.

We detected significant departures from neutrality expectations (table 1), as measured by Tajima's D, in the global data set and the pooled African and non-African data sets, but not in Tanzanians. D^* and F^* statistics of Fu and Li revealed significant departures from neutrality in all populations. We also tested all mtDNAs belonging to subsets of haplogroups L0 and L1 for deviations from neutrality expectations (results not shown). None of these subsets significantly deviated from neutrality expectations, except for mtDNAs belonging to L5 (Tajima's $D = -1.29$, $P < 0.001$; D^* of Fu and Li $= -1.26$, $P < 0.05$; F^* of Fu and Li $= -1.37$, $P < 0.05$). Other studies of whole-mtDNA genome diversity in Africa did not report significant deviations from neutrality expectations (Ingman et al. 2000; Mishmar et al. 2003; Ruiz-Pesini et al. 2004), with the exceptions of Torroni et al. (2001) and Kivisild, Shen, et al. (2006). We speculate that these departures from neutrality expectations may be attributed to several potential factors including our larger sample of African mtDNAs compared with previous studies, artifacts of sampling from structured populations (resulting in an excess of rare variants) (Ptak and Przeworski 2002; Hammer et al. 2004; Kivisild, Shen, et al. 2006), population expansion events, and/or an ancient selective sweep in African mtDNAs (Simonsen et al. 1995).

Due to the high frequency of homoplasy in the mtDNA d-loop, we compared the topology of an NJ tree reconstructed from the complete mtDNA sequences of the 322 samples (fig. S1, Supplementary Material online) with a tree reconstructed using the mtDNA sequences excluding the d-loop (fig. 2). The topologies of the 2 trees were similar, but the basal branches of the complete mtDNA sequences had lower bootstrap values. In contrast, the NJ tree of the mtDNA genomes that excluded the d-loop had higher statistical support for the basal branches separating the haplogroups (L0, L1, L5, L2, L3, M, and N), with bootstrap values ranging from 61% to 91%.

There are several notable characteristics of the NJ tree shown in figure 2. First, the L0/L5/L1/L2/L3 haplogroups are African specific, as previously reported (Ingman et al. 2000; Mishmar et al. 2003; Ruiz-Pesini et al. 2004; Kivisild, Shen, et al. 2006). Haplogroup L0 forms the basal lineage of the human mtDNA gene tree followed by L1, L5, L2, L3, M, N, and the derivatives of haplogroups M and N. Unlike previous studies that report L0k as the most basal haplogroup of the human mtDNA gene tree (Ingman et al. 2000; Mishmar et al. 2003; Ruiz-Pesini et al. 2004), our data suggest that L0d forms the basal lineage of the human mtDNA gene tree (which has also been observed in a recent study of African mtDNA genome diversity [Kivisild, Shen, et al. 2006]). L0d is further subdivided into 2 reciprocally monophyletic clades: one clade composed of SAK and one clade composed of Tanzanians (all of whom are Sandawe, except for one neighboring Burunge). Further, our analysis shows that L0k forms a clade with L0f and L0a, suggesting independent origins of L0d and L0k lineages. L1b, L1c, and L5 form a monophyletic clade, as do all L2 mtDNAs and all L3 lineages. These L3 mtDNA sequences all belong to Africans, with the exception of 2 mtDNA genomes sampled from the Middle East, suggesting some recent migration from Africa to the Middle East. mtDNAs belonging to haplogroups M and N form 2 monophyletic clades (fig. 2*A*). These 2 M and N haplogroup clades included a few Tanzanians (belonging to haplogroups M1, M, N1, and J), suggesting possible recent gene flow back into Africa and/or that ancestors of the Tanzanian populations may have been a source of migration of modern humans from Africa to other regions (fig. 2*B*).

In order to better resolve the evolutionary history of the most ancient mtDNA haplogroup lineages using Bayesian maximum likelihood analyses, we next analyzed a smaller data set composed of all mtDNA genomes of people from sub-Saharan Africa and a subset of the samples obtained from GenBank. The subset of samples included a global panel representing all of the major non-African haplogroup lineages. The Bayesian tree is shown in figure 3. The overall tree topologies of the Bayesian tree and NJ tree were similar. Clade credibility scores, which are a measure of the posterior probability of the tree branching structure, ranged from 73% to 100%. MtDNAs of Africans belonging to haplogroups L0 and L1 form the most basal lineages of the human mtDNA gene tree. Within L0, L0d forms the most basal branch of the tree and also contains 2 reciprocally monophyletic clades composed of Tanzanians and SAK, respectively. L0k forms a clade with L0f and L0a, providing additional support of independent origins of the Khoisan-specific L0d and L0k haplogroup lineages. L1b and L1c form a clade that does not include L5. In contrast to the NJ trees (fig. 2 and fig. S1, Supplementary Material online), L5 occupies an intermediate phylogenetic position between L1 and L2, as has been previously reported (Shen et al. 2004).

Using a log likelihood test (Huelsenbeck and Crandall 1997), we found significant substitution rate heterogeneity in our phylogenetic analyses for the complete sample of mtDNA genomes ($n = 320$; $-2ln\Lambda = 750.9$, χ^2 df $= 318$, $P < 0.05$, 100 permutations) and for the smaller data set ($n = 114$; $-2ln\Lambda = 349.5$, χ^2 df $= 112$, $P < 0.05$, 100 permutations). Simulations have shown that it is difficult to root a phylogeny precisely when the outgroup is very distant relative to the ingroup, as is the case in the present study (Penny et al. 1995; Sanderson and Shaffer 2002). This difficulty could potentially explain the substitution rate heterogeneity observed in our data set. Reanalysis of the data set using a conservative midpoint root verified that the substitution rates in our data were heterogeneous for the complete set of mtDNA genomes ($n = 319$; $-2ln\Lambda = 433.94$, χ^2 df $= 317$, $P < 0.05$, 100 permutations) and for the smaller data set ($n = 113$; $-2ln\Lambda = 150.41$, χ^2 df $= 111$, $P < 0.05$, 100 permutations).

Subsequent to our discovery that these data do not follow a clock-like model, we applied a PL algorithm to account for substitution rate heterogeneity among the mtDNA haplogroup clades to calculate TMRCAs for various nodes in the gene tree shown in figure 3. Table 2 lists these TMRCA dates and their 95% CIs. Our TMRCA estimate for the global mtDNA genome tree is 194.3 ± 32.55 kya, which is very close to the age of the earliest modern humans estimated from fossil data (McDougall et al. 2005) as well as some early studies of mtDNA diversity (e.g., Vigilant et al. 1991; Horai et al. 1995 [when corrected for a *Pan*/*Homo*

FIG. 2.—NJ trees based on mtDNA genomes (excluding the d-loop). (*A*) Samples are colored according to their haplogroup membership. (*B*) Samples are colored according to their geographic origin (sub-Saharan Africa, North Africa/Middle East/Eurasia, Asia [northern, central, and eastern], Southeast Asia/Oceania, and Americas). Samples from AF381984 (Morroco) and AF381996 (Jordan) that belong to African L3 haplogroups indicated by "+." Samples of Tanzanians belonging to haplogroup M are indicated by "*." Samples of Tanzanians belonging to haplogroup N are indicated by "**."

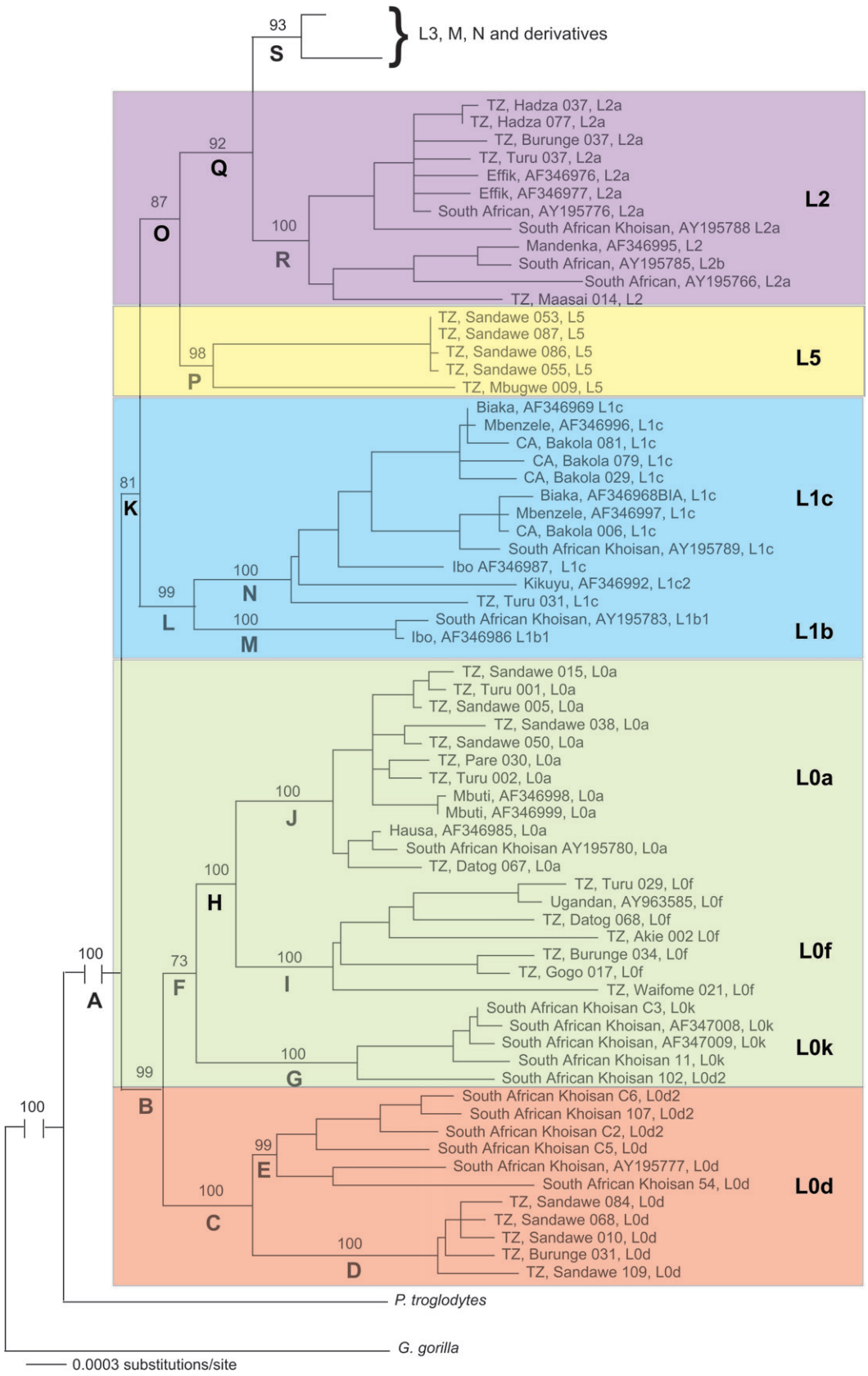

FIG. 3.—Phylogenetic tree of mtDNA genomes (excluding the d-loop) obtained by maximum likelihood Bayesian analysis.

Table 2
TMRCA Dates Inferred from Tree Nodes Shown in Figure 3

Node	Lineage	MRCA ± CI × 10^3 years
	Pan/Homo	6,500.0[a]
A	*Homo sapiens sapiens*	194.3 ± 32.5
B	L0	146.4 ± 25.1
C	L0d	106 ± 20.2
D	Tanzanian L0d	30.6 ± 17.8
E	San L0d	90.4 ± 18.9
F	L0k, L0f, L0a	139.8 ± 24.6
G	L0k	70.9 ± 19.7
H	L0f, L0a	100.1 ± 12.5
I	L0f	94.9 ± 9.4
J	L0a	54.6 ± 5.7
K	L1, L2, L3, M, N	142.3 ± 38.2
L	L1b, L1c	140.4 ± 32.9
M	L1b	15 ± 26.7
N	L1c	95.6 ± 23.9
O	L5, L2, L3, M, N	131.2 ± 15.9
P	L5	129.4 ± 22.1
Q	L2, L3, M, N	96.7 ± 10.7
R	L2	94.5 ± 4.5
S	L3, M, N, and derivatives	94.3 ± 9.9

[a] We assumed a *Pan/Homo* divergence of 6.5 Myr.

split 6.5 MYA]). We also observe an origin of L0 (146.4 ± 25.1 kya) and L1 (140.4 ± 33 kya), slightly more recent than the appearance of modern humans based on the paleontological record (Clark et al. 2003; White et al. 2003; McDougall et al. 2005). The L0d mtDNAs have a TMRCA of 106 ± 20.2 kya. The TMRCA of mtDNAs of the SAK belonging to L0d is 90.4 ± 18.9 kya, whereas the TMRCA of L0d mtDNAs belonging to Tanzanians is more recent (30.6 ± 17.8 kya). The TMRCA of L0k, L0f, and L0a is 139.8 ± 24.6 kya. The TMRCA of the SAK L0k is 70.9 ± 19.7 kya. The TMRCA of L0f, which is observed only in eastern Africa, indicates that it is a relatively old lineage (94.9 ± 9.4 kya). The TMRCA of L0a (54.6 ± 5.7 kya) is more recent than the TMRCA of L0f, even though these mtDNA samples originate from diverse regions in Africa. We attribute the relatively old TMRCA (and highly negative Tajima's *D*) of L5 (129.4 ± 22.1 kya) to the divergent sequence of the L5 mtDNA from a single Tanzanian Mbugwe individual compared with the three L5 mtDNAs from the Tanzanian Sandawe that differed from each other by very few basepairs. The TMRCAs of L2 and L3 are more recent (96.7 ± 10.7 kya) compared with those of L0, L1, and L5. The age of the youngest node containing both African and non-African sequences (node S) is 94.3 ± 9.9 kya and represents an upper bound time estimate for an exodus out of Africa.

Phylogenetic analyses of mtDNA that assume a strict bifurcating tree topology may not be well suited to the study of human mtDNA (Bandelt et al. 1999). Confounding phenomena, such as homoplasy and multifurcations ("star" phylogenies), have commonly been observed in studies of human mtDNA gene genealogies (Posada and Crandall 2001). Therefore, we complemented our phylogenetic analyses by also constructing mtDNA gene genealogies (fig. S2, Supplementary Material online) using a MJ network approach. Generally, the results of our network analysis were consistent with haplogroup designations based on d-loop sequences and SNP analysis reported in other studies (Salas et al. 2002, 2004; Kivisild et al. 2004). Using the network analysis, we investigated the substitutions in the mtDNA coding region that occur along main branches of the network (fig. 4). Additional substitutions that occur along the terminal branches of the network are given in figures S3 and S4 (Supplementary Material online). The main branches of the L0/L1 network are well supported by several substitutions across the mtDNA genome. Note that L0d, L0f, L0k, and L1c haplogroups contain highly divergent lineages that are separated by several substitutions. In contrast, the L0a haplogroup was characterized by several short branches, despite the fact that these mtDNAs come from a wide variety of African populations. We speculate that the wide geographic distribution and short branches of L0a result from recent population growth and/or a recent expansion of L0a and could perhaps reflect the expansion of Bantu-speaking peoples into eastern Africa within the past few thousand years (Salas et al. 2002).

Discussion

Most analyses of the phylogenetic relationships among African mtDNA haplogroup lineages have been confined to the d-loop and/or RFLP haplotyping of the whole-mtDNA genome. Phylogenies and TMRCA estimates based on the d-loop and RFLPs may be problematic because of homoplasy and heterogeneous mutation rates (Maddison et al. 1992; Excoffier and Yang 1999; Meyer et al. 1999; Ballard and Whitlock 2004). Although previous studies of whole-mtDNA genome diversity have included over 100 mtDNA genomes of Africans (Ruiz-Pesini et al. 2004; Kivisild, Shen, et al. 2006), these studies have not included large numbers of samples from eastern African populations. However, eastern African populations contain rare mtDNA haplogroups that may contain important clues in understanding modern human origins. Our analysis of mtDNA genomes provides relatively robust phylogenies and TMRCA estimates for these mtDNA haplogroup lineages. Moreover, the results of our study suggest several notable observations about the role of Tanzanians in the dispersion of modern humans and the history of African mtDNA haplogroups.

Tanzania is the only region of Africa where populations speak languages classified as belonging to the 4 major language families present in Africa: Afro-Asiatic, Nilo-Saharan, Niger-Kordofanian, and Khoisan (Greenberg 1963). The Hadza and Sandawe, who speak a click language classified as Khoisan, are thought to be indigenous to Tanzania. However, populations speaking languages belonging to the other 3 language families are thought to have migrated into Tanzania from the Sudan (Nilotic Nilo-Saharan speakers), Ethiopia (Cushitic Afro-Asiatic speakers), and West Africa (Bantu Niger-Kordofanian speakers) within the past 5,000 years (Ambrose 1982; Newman 1995). Given the considerable ethnic and linguistic diversity present in Tanzania, it is not surprising that Tanzanians possess high mtDNA genetic diversity, comparable to the level of genetic diversity observed across continental sub-Saharan Africa. This genetic diversity is distributed among several mtDNA haplogroups that originated at different times in modern human history. The presence of very old mtDNA haplogroups (i.e., L0d, L0f, and L5) in Tanzanians that are

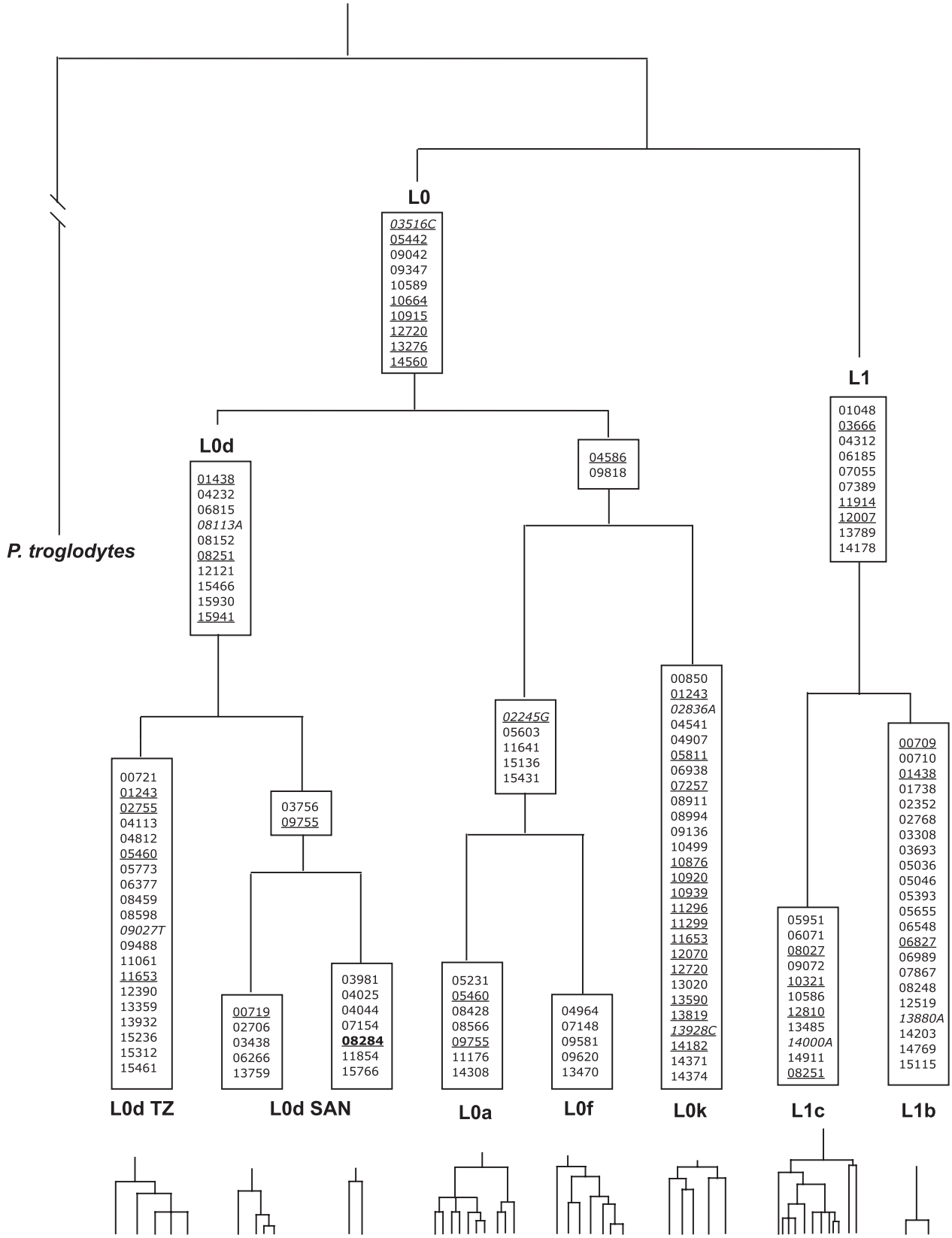

FIG. 4.—Phylogenetic relationships of haplogroups L0 and L1 mtDNA genomes (excluding the d-loop) inferred from a reduced median network analysis using Network4.111. *Pan troglodytes* was included as an outgroup to root the network. Nucleotide substitutions supporting each branch are shown along the phylogeny. Recurrent mutations are underlined. Transversions are shown in italic, with the nucleotide change indicated. Indels are indicated with boldface font. Additional information about substitutions that define the terminal nodes are given in figures S2–S4 (Supplementary Material online). Numbers refer to their position relative in the Cambridge Reference Sequence (Andrews et al. 1999).

rare or absent in other regions of Africa suggests populations in Tanzania may have had a large long-term effective population size and/or a large degree of long-term population structure, which has acted to preserve many divergent and rare mtDNA haplogroup lineages that appeared early in modern human history. The presence of these ancient lineages in Tanzania also suggests that eastern Africa might be the source of origin of many other African mtDNA haplogroup lineages. Our findings are consistent with other studies of mtDNA genetic diversity in African populations that have suggested populations in eastern Africa form a highly diverse gene pool (Watson et al. 1997; Chen et al. 2000; Watson and Penny 2003; Kivisild et al. 2004). In addition, the TMRCA of mtDNA haplogroup lineages L3, M, and N and their derivatives (94.3 $\pm$ 9.9 kya) is approximately half of the TMRCA of all modern humans (194.3 $\pm$ 32.55 kya), which supports models predicting that there was a significant period of time in which modern humans lived exclusively in Africa prior to the exodus of modern humans to other regions of the world (Penny et al. 1995). These observations are consistent with paleobiological and archeological data suggesting that eastern Africa may have been an ancient source of dispersion both within and outside of Africa. The earliest remains of transitional modern humans, dated as early as 196 kya, have been found in Ethiopia (Clark et al. 2003; White et al. 2003; McDougall et al. 2005). The earliest artifacts associated with modern humans are also found in eastern Africa (Foley 1998). Later, Stone Age technology was established in several regions well before 40 kya in eastern Africa but not until 22 kya in southern Africa (Lahr and Foley 1994; Lahr 1996; Foley 1998).

Further, the reciprocally monophyletic phylogenetic relationship of L0d lineages in the Sandawe and the SAK at the root of the human mtDNA gene tree, indicates an ancient, but unique, genetic connection between these populations (Tishkoff SA and Mountain JL, unpublished data). The oldest L0d lineages are observed in the SAK, but it is possible that the ancestral Khoisan population(s) originated in east Africa and subsequently migrated into southern Africa, and that ancient lineages have been lost in the Tanzanian Hadza and Sandawe populations due to genetic drift (Tishkoff SA and Mountain JL, unpublished data). These observations are consistent with both linguistic data indicating similarities between the Sandawe and SAK languages (Ruhlen 1991; Ehret 2000; Traunmuller 2003) as well as shared subsistence patterns (until recently, the Sandawe maintained a hunter–gatherer lifestyle). Our findings are also consistent with patterns of variation in the Y chromosome suggesting an ancient genetic connection between SAK and several East African populations (Cruciani et al. 2002; Semino et al. 2002). Additional data from other loci and additional populations from Tanzania will help resolve whether the connection between Khoisan speakers in eastern and southern Africa is due to divergence from a common ancestor, or to ancient gene flow, and whether or not the ancestors of the Khoisan-speaking populations originated in eastern or southern Africa (Tishkoff SA and Mountain JL, unpublished data).

Finally, our limited genetic data from Tanzanians belonging to haplogroups M1, N1, and J suggest 2 alternatives that are not mutually exclusive. Populations in Tanzania may have been important in the migration of modern humans from Africa to other regions, as noted in previous studies of other populations in eastern Africa (Quintana-Murci et al. 1999). For example, mtDNAs of Tanzanians belonging to haplogroup M1 cluster with peoples from Oceania, whereas Tanzanian mtDNAs belonging to haplogroup N1 and J cluster with peoples of Middle Eastern and Eurasian origin. However, the presence of haplogroups N1 and J in Tanzania suggest "back" migration from the Middle East or Eurasia into eastern Africa, which has been inferred from previous studies of other populations in eastern Africa (Kivisild et al. 2004). These results are intriguing and suggest that the role of Tanzanians in the migration of modern humans within and out of Africa should be analyzed in greater detail after more extensive data collection, particularly from analysis of Y-, X-, and autosomal chromosome markers. Our analyses of African mtDNAs suggest populations in eastern Africa have played an important and persistent role in the origin and diversification of modern humans.

Supplementary Material

Supplementary figures S1–S4 and tables S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This study was funded by L.S.B. Leakey Foundation, Wenner Gren Foundation, NSF BCS-0196183, BSC-0552486, and Packard and Burroughs Wellcome Foundation grants to S.A.T. H.M.M. and A.de S. were funded by IGERT-9987590 grant to S.A.T. F.A.R. was supported by National Institutes of Health grant F32HG003801. We thank Kweli Powell, Maritha Koetze and Alain Froment for assistance with DNA sample collection and Nigel Crawhall, Christopher Ehret, Alison Brooks, and Joanna Mountain for helpful discussion. We thank Godfrey Lema, Salum Juma Deo, Paschal Lufungulo, Waja Ntandu, Dr T.B. Nyambo at MUCHS, Dr Audax Mabulla at University of Dar es Saalam, Jeannette Hanby, and David Bygott for their assistance with field work in Tanzania. We thank African participants who generously donated DNA samples so that we might learn more about their population history.

Literature Cited

Ambrose SH. 1982. Archaeology and linguistic reconstructions of history in eastern Africa. In: Ehret C, Posnansy M, editors. Archaeological and linguistic reconstruction of African history. Berkeley (CA): University of California Press. p. 104–157.

Andrews R, Kubacka I, Chinnery P, Lightowlers R, Turnbull D, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 2:149.

Baldwin BG, Sanderson MJ. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc Natl Acad Sci USA. 95:9402–9406.

Ballard JW, Whitlock MC. 2004. The incomplete natural history of mitochondria. Mol Ecol. 13:729–744.

Bandelt H, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.
Bandelt HJ, Kong QP, Richards M, Villems R, Macaulay V. 2006. Estimation of mutation rates and coalescence times: some caveats. In: Bandelt HJ, Macaulay V, Richards M, editors. Human mitochondrial DNA and the evolution of *Homo sapiens*. Berlin (Germany): Springer-Verlag. p. 149–179.
Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. Nature. 325:31–36.
Chen J, Sokal RR, Ruhlen M. 1995. Worldwide analysis of genetic and linguistic relationships of human populations. Hum Biol. 67:595–612.
Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC. 2000. mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. Am J Hum Genet. 66:1362–1383.
Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet. 57:133–149.
Clark JD, Beyene Y, WoldeGabriel G, et al. (13 co-authors). 2003. Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. Nature. 423:747–752.
Cruciani F, Santolamazza P, Shen P, et al. (16 co-authors). 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet. 70:1197–1214.
Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglia A, Pascali V, Spedini G, Calafell F. 2004. The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. Am Nat. 163:212–226.
Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19:4008.
Ehret C. 2000. Language and history. In: Heine B, Nurse D, editors. African languages: an introduction. Cambridge (UK): Cambridge University Press.
Excoffier L, Yang Z. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol. 16:1357–1368.
Felenstein J. 1993. PHYLIP: phylogenetic inference package. Seattle (WA): Department of Genetics, University of Washington.
Finnila S, Lehtonen MS, Majamaa K. 2001. Phylogenetic network for European mtDNA. Am J Hum Genet. 68:1475–1484.
Foley R. 1998. The context of human genetic evolution. Genome Res. 8:339–347.
Greenberg J. 1963. The languages of Africa. Bloomington (IN): Indiana University Press.
Hammer MF, Garrigan D, Wood E, Wilder JA, Mobasher Z, Bigham A, Krenz JG, Nachman MW. 2004. Heterogeneous patterns of variation among multiple human X-linked loci: the possible role of diversity-reducing selection in non-Africans. Genetics. 167:1841–1853.
Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA. 92:532–536.
Huelsenbeck JP, Crandall KA. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst. 28:437–466.
Ingman M, Gyllensten U. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res. 13:1600–1606.
Ingman M, Kaessmann H, Paabo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature. 408:708–713.
Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet. 66:979–988.
Kivisild T, Metspalu M, Bandelt HJ, Richards M, Villems R. 2006. The world mtDNA phylogeny. In: Bandelt HJ, Macaulay V, Richards M, editors. Human mitochondrial DNA and the evolution of *Homo sapiens*. Berlin (Germany): Springer-Verlag. p. 149–179.
Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet. 75:752–770.
Kivisild T, Shen P, Wall DP, et al. (17 co-authors). 2006. The role of selection in the evolution of human mitochondrial genomes. Genetics. 172:373–387.
Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. Curr Biol. 13:464–473.
Kumar S, Filipski A, Swarna V, Walker A, Hedges SB. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. Proc Natl Acad Sci USA. 102:18842–18847.
Lahr MM. 1996. The evolution of modern human diversity. Cambridge (UK): Cambridge University Press.
Lahr MM, Foley RA. 1994. Multiple dispersals and modern human origins. Evol Anthropol. 3:48–60.
Macaulay V, Hill C, Achilli A, et al. (21 co-authors). 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science. 308:1034–1036.
Maddison DR, Ruvolo M, Swofford DL. 1992. Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. Syst Biol. 41:111–124.
Maddison WP, Maddison DR. 2000. MacClade: analysis of phylogeny and character evolution. Sunderland (MA): Sinauer Associates, Inc.
McBrearty S, Brooks A. 2000. The revolution that wasn't: a new interpretation of the origin of modern human behavior. J Hum Evol. 39:453–563.
McDougall I, Brown FH, Fleagle JG. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature. 433:733–736.
Meyer S, Weiss G, von Haesler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics. 152:1103–1110.
Mishmar D, Ruiz-Pesini E, Brandon M, Wallace DC. 2004. Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. Hum Mutat. 23:125–133.
Mishmar D, Ruiz-Pesini E, Golik P, et al. (13 co-authors). 2003. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA. 100:171–176.
Newman J. 1995. The peopling of Africa. New Haven (CT): Yale University Press.
Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature. 441:1103–1108.
Penny D, Steel M, Waddell PJ, Hendy MD. 1995. Improved analyses of human mtDNA sequences support a recent African origin for Homo sapiens. Mol Biol Evol. 12:863–882.
Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet. 65:439–458.
Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol. 16:37–45.

Ptak SE, Przeworski M. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet. 18:559–563.
Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS. 1999. Genetic evidence of an early exit of *Homo sapiens* sapiens from Africa through eastern Africa. Nat Genet. 23:437–441.
Rieder MJ, Taylor SL, Tobe VO, Nickerson DA. 1998. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Res. 26:967–973.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572–1574.
Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R. 2004. MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. Ann Hum Genet. 68: 340–352.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
Ruhlen MA. 1991. Guide to the world's languages. Stanford (CA): Stanford University Press.
Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. Science. 303:223–226.
Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet. 71:1082–1111.
Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet. 74:454–465.
Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol. 14:1218–1231.
Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 19:101–109.
Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19:301–302.
Sanderson MJ, Doyle JA. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. Am J Bot. 88:1499–1516.
Sanderson MJ, Shaffer HB. 2002. Troubleshooting molecular phylogenetic analyses. Annu Rev Ecol Syst. 33:49–72.
Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA. 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am J Hum Genet. 70:265–268.
Shen P, Lavi T, Kivisild T, et al. (11 co-authors). 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum Mutat. 24:248–260.
Simonsen KL, Churchill GA, Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics. 141:413–429.
Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud RP, Spadoni JL, El-Chenawi F, Beraud-Colomb E. 2004. Mitochondrial DNA sequence diversity in a sedentary population from Egypt. Ann Hum Genet. 68:23–39.
Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony. Sunderland (MA): Sinaeur Associates, Inc.
Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10:512–526.
Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L. 2005. Reconstructing the origin of Andaman Islanders. Science. 308:996.
Thompson JD, Gibson TJ, Plewniak AJ, Kidd JR. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24:4876–4882.
Tishkoff SA, Dietzsch E, Speed W, et al. (11 co-authors). 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science. 271:1380–1387.
Torroni A, Rengo C, Guida V, et al. (12 co-authors). 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet. 69:1348–1356.
Traunmuller H. 2003. Clicks and the idea of a human protolanguage. Phonum. 9:1–4.
Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet. 65:43–62.
Underhill PA, Shen P, Lin AA, et al. (21 co-authors). 2000. Y chromosome sequence variation and the history of human populations. Nat Genet. 26:358–361.
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science. 253:1503–1507.
Wallace DC, Brown MD, Lott MT. 1999. Mitochondrial DNA variation in human evolution and disease. Gene. 238:211–230.
Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Paabo S. 1996. mtDNA sequence diversity in Africa. Am J Hum Genet. 59:437–444.
Watson E, Forster P, Richards M, Bandelt HJ. 1997. Mitochondrial footprints of human expansions in Africa. Am J Hum Genet. 61:691–704.
Watson EE, Penny D. 2003. DNA and people: from East Africa to Aotearoa New Zealand. N Z Sci Rev. 60:29–35.
White T, Asfaw B, DeGusta D, Gilbert H, Richards G, Suwa G, Howell F. 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. Nature. 423:742–747.

Lisa Matisoo-Smith, Associate Editor

Accepted December 12, 2006