# Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears

Toomas Kivisild,[1] Maere Reidla,[1] Ene Metspalu,[1] Alexandra Rosa,[1] Antonio Brehm,[2] Erwan Pennarun,[1] Jüri Parik,[1] Tarekegn Geberhiwot,[3] Esien Usanga,[4] and Richard Villems[1]

[1]Estonian Biocentre and Tartu University, Tartu, Estonia; [2]Center of Macaronesian Studies, University of Madeira Penteada, Funchal, Portugal; [3]Birmingham and Solihull Teaching Hospital, Birmingham, United Kingdom; and [4]Department of Haematology, University of Calabar, Calabar, Nigeria

**Approximately 10 miles separate the Horn of Africa from the Arabian Peninsula at Bab-el-Mandeb (the Gate of Tears). Both historic and archaeological evidence indicate tight cultural connections, over millennia, between these two regions. High-resolution phylogenetic analysis of 270 Ethiopian and 115 Yemeni mitochondrial DNAs was performed in a worldwide context, to explore gene flow across the Red and Arabian Seas. Nine distinct subclades, including three newly defined ones, were found to characterize entirely the variation of Ethiopian and Yemeni L3 lineages. Both Ethiopians and Yemenis contain an almost-equal proportion of Eurasian-specific M and N and African-specific lineages and therefore cluster together in a multidimensional scaling plot between Near Eastern and sub-Saharan African populations. Phylogeographic identification of potential founder haplotypes revealed that approximately one-half of haplogroup L0–L5 lineages in Yemenis have close or matching counterparts in southeastern Africans, compared with a minor share in Ethiopians. Newly defined clade L6, the most frequent haplogroup in Yemenis, showed no close matches among 3,000 African samples. These results highlight the complexity of Ethiopian and Yemeni genetic heritage and are consistent with the introduction of maternal lineages into the South Arabian gene pool from different source populations of East Africa. A high proportion of Ethiopian lineages, significantly more abundant in the northeast of that country, trace their western Eurasian origin in haplogroup N through assorted gene flow at different times and involving different source populations.**

## Introduction

East Africa occupies a central position in the emergence of hominid species, since it is the location of the earliest fossil evidence for anatomically modern humans, dating back 150,000–160,000 years (Clark et al. 2003; White et al. 2003). Its geographic position makes it one of the most likely sections of Africa from which the colonization of Eurasia by the ancestors of European, Asian, and Oceanian populations started ~50,000–70,000 years ago: both the "southern" and "northern" routes can be logically drawn as springing from there (Sauer 1962; Cavalli-Sforza et al. 1994; Lahr and Foley 1994; Stringer 2000; Walter et al. 2000; Kivisild et al. 2003a). It is unfortunate that archaeological data for the late Pleistocene and early Holocene in Ethiopia is limited. Obsidian-trading links between coastal East Africa and Arabia can be traced back to the 7th millennium B.C., whereas the beginning of agriculture in Ethiopia is usually attributed to increasing contacts with Egypt and the Near East,

from the middle of the 5th millennium B.C. Records concerning the trade of myrrh between Egypt and Ethiopia, along the Red Sea coast, go back to the 3rd millennium B.C. The Horn of Africa may have been the major prehistoric entry point of the African zebu cattle from the southern Arabian peninsula (Hanotte et al. 2002). In the late 2nd and early 1st millennium B.C., the eastern part of the Tigray plateau was included in the Tihama cultural complex, spread widely along both the African and Arabian coasts of the Red Sea (Munro-Hay 1991). Later, in the mid-1st millennium B.C., the southern Arabian immigrants, or Sabaens, appeared in Ethiopia. These were military or trading colonists who maintained contact with their country of origin for centuries. The Semitic-speaking Aksumites, or Habash (Abyssinians), had their capital city, Aksum, in the western part of the province of Tigray. During the first 6 centuries A.D., they controlled territories north to upper Egypt, east to the Gulf of Aden and southern Arabia, south to the Omo River, and west to the Cushite Kingdom of Meroë (Munro-Hay 1991). The ethnonym "Ethiopians"—the people with the "burnt face"—was coined by the Greeks, although it may originally have been applied to the Nubians, who were part of the Cushite kingdom. The Arab slave trade in the Indian Ocean was less intense and more sporadic than the slave trade on the Atlantic coast of

Africa. However, during the 8th, 9th, and 19th centuries, it is estimated that millions of East Africans were deported, the majority to the Arabian Peninsula, Near Eastern countries, and India. The main source regions for slaves in East Africa extended from the interior of present-day Mozambique to Ethiopia (Harris 1971).

The three major ethnic groups of Ethiopia today are the Tigrais, Amharas, and Oromos. Together, they account for approximately three-quarters of the total national population. Amharas, Tigrais, and Gurages speak Semitic languages and are considered to be descendants of southern Arabian conquerors, who trace their ancestry back to Moses and King Solomon. Whereas Tigrais still live in the area of the ancient Aksum kingdom, the Amharas and Gurages have expanded inland. Because Amharas have largely taken the role of the political and cultural elite in the country, there is a process of "amharization," which can be understood, at least partly, as a matter of prestige and which leads to the cultural assimilation of other minority populations. The Oromos and Afars speak Cushitic languages and are purported to have connections to ancient Egyptians, since the land of Cush—the son of biblical Ham—is generally considered to be in the vicinity of the ancient cities of Meroë and Napata, located in present-day Sudan. Yet it should be stressed here that the split between the Cushitic and Semitic languages, branches of the Afro-Asiatic linguistic family, is ancient, probably predating the Holocene (see, e.g., Militarev [2003]).

The linguistic reconstructions of Semitic vocabulary, related to farming and agriculture, have supported the theory that the origin of Semitic languages is in the Near East (Diakonoff 1988; Militarev 2003). On the other hand, the finding of all major branches of the Afro-Asiatic language tree in Ethiopia, including those that are not spoken elsewhere in the world, suggests that the homeland of the Afro-Asiatic language family may have been somewhere close to southwestern Ethiopia (Ehret 1995). However, both cultural and historic evidence show tight connections between East Africa and the Semitic cultural substrate in the Near East and southern Arabia, which points to four distinct phases of Semitic cultural intrusion into Ethiopia: first, related to the Sabaens in the 1st millennium B.C.; second, as the arrival of Falasha Jews from southern Arabia in the first 2 centuries A.D.; third, during the 4th–6th centuries, when Syrian missionaries brought Christianity to Aksumites and to their descendants, the Tigrais and the Amharas; and fourth, because of the influence of Muslim Arabs, which primarily affected the southeastern parts of the country (Levine 1974).

Previous studies of classic genetic markers and Y-chromosomal haplogroup distributions have shown that, in addition to the predominant sub-Saharan African substrate, the Ethiopian gene pool also embraces a considerable component indicative of admixture with populations of Arabian and/or Near Eastern origin (Cavalli-Sforza 1997; Passarino et al. 1998; Thomas et al. 2000; Cruciani et al. 2004; Luis et al. 2004). The current classification of African mtDNA lineages recognizes haplogroups L0–L3 and L5 as the five major mtDNA haplogroups whose spread is restricted mainly to sub-Saharan Africa (Rosa et al. 2004; Salas et al. 2004; Shen et al. 2004). The regional subclusters of these haplogroups are often associated with high sequence variation (Vigilant et al. 1991; Chen et al. 1995, 2000; Graven et al. 1995; Watson et al. 1997; Rando et al. 1998; Krings et al. 1999; Salas et al. 2002; Rosa et al. 2004). Although a number of complete sequences of African origin are available—largely from African Americans, representative of genetic variation mostly in West Africa—the regional variants from East Africa are still incompletely characterized (Ingman et al. 2000; Maca-Meyer et al. 2001; Herrnstadt et al. 2002; Mishmar et al. 2003; Howell et al. 2004). Two previous studies of mtDNA variation in Ethiopia (Passarino et al. 1998; Quintana-Murci et al. 1999) found that a substantial proportion of lineages derive from haplogroups M and N. On the other hand, it has been proposed that, in contrast to Y-chromosome lineages, a substantial proportion of Arabian mtDNA lineages trace their origin to East Africa, which suggests that the female slaves from Africa made a relatively major long-term contribution to the gene pool of southern Arabians (Richards et al. 2003). We characterize, at high phylogenetic resolution, an extraordinary level of mtDNA variation in different Ethiopian populations, to infer the role of the region in the genetic history of anatomically modern humans. A parallel study of Yemeni populations was performed to assess the levels of gene flow between East Africa and Arabia. On the basis of a detailed analysis of both control- and coding-region variation and by comparison of the results with data available from other African and Near Eastern populations, we attempted to distinguish which regions of East Africa have contributed to the gene flow to southern Arabia and to identify possible source areas of western Eurasian haplogroups that are present in Ethiopians.

## Material and Methods

DNA samples of 53 Tigrais from Ethiopia and Eritrea (Semitic-language speakers), 120 Amharas (Semitic), 33 Oromos (Cushitic), 21 Gurages (Semitic), 16 Afars (Cushitic), and 28 individuals with other ethnic affiliations were obtained, in five different cities of Ethiopia (fig. 1), from maternally unrelated volunteers, all of whom gave informed consent. In addition, blood was drawn, after informed consent was obtained, from 115 volunteer Yemeni donors in Kuwait who claimed their maternal

**Figure 1**    Map of East Africa and Arabia, with locations of samples collected for the present study. The five cities in Ethiopia where DNA samples were collected are indicated. The main spread zones of the five major Ethiopian linguistic groups that are discussed in the study are shown in blue capital letters.

origin was in Yemen. The first hypervariable segment (HVS-I) was sequenced in all DNA samples (EMBL Nucleotide Sequence Database [accession numbers AJ748863–AJ749247]) and informative coding-region markers (Ingman et al. 2000; Torroni et al. 2001; Herrnstadt et al. 2002; Maca-Meyer et al. 2003; Mishmar et al. 2003), and HVS-II sequences were determined in selected individual samples by use of RFLP assays or direct sequencing. Classification of subclades of haplogroups L0, L1, L2, and L3 follows the nomenclature of Torroni et al. (2001), Salas et al. (2004), and Shen et al. (2004). Where coding region resequencing revealed novel parsimony, informative polymorphisms, or previously unidentified splits in the inner branches of the mtDNA phylogeny, new haplogroups were defined that extended the framework of the existing classification scheme.

Phylogenetic networks were drawn by hand, according to the guidelines of the median-joining algorithm (Bandelt et al. 1999), for each haplogroup independently. Subsequently, the most parsimonious tree of haplogroups was inferred. Coalescent times were calculated using the $\rho$ statistic and an HVS-I–mutation rate of one transition per 20,180 years in base pairs 16090–16365 (Forster et al. 1996; Saillard et al. 2000). Nei's haplotype diversities were calculated using the formula $h = [n/(n-1)](1 - p_i^2)$, where $n$ is the sample size and $p_i$ the frequency of individual haplotypes in that sample (Nei 1987). Pairwise $F_{ST}$ distances between populations were calculated from haplogroup frequencies, and their significance was tested with a nonparametric permutation test by use of the ARLEQUIN package (Schneider et al. 2000). Multidimensional scaling (MDS) plots were obtained with SPSS 10. Admixture proportions were calculated with the ADMIX 2.0 program (Dupanloup and Bertorelle 2001).

## Results

A total of 168 different mtDNA haplotypes were observed in 270 Ethiopians and Eritreans, and 72 haplotypes were recovered in 115 Yemeni samples (fig. 2; tables A1–A6 [online only]). Approximately one-half of both Ethiopian (52.2%) and Yemeni (45.7%) mtDNA lineages belonged to clades specific to sub-Saharan Africa (fig. 2A; table 1), whereas the other half was divided between derived subclades of haplogroups M and N (fig. 2B; table 1) that are, with the exception of M1 and U6 lineages, more common outside Africa. Consistent with the coexistence of sub-Saharan African and Eurasian mtDNA lineages among Ethiopian, Egyptian, and Yemeni populations, the MDS plot (fig. 3) clustered them, together with Egyptians, in between the Near Eastern and the West African and southern African clusters. It is interesting that both Semitic- and Cushitic-speaking populations of Ethiopia were close to each other and

did not reveal significant differences ($P > .05$) in $F_{ST}$ distances between themselves (table A7). The differences between Ethiopian and Yemeni populations were significant ($P < .01$) except in the case of Gurages ($P = .0992 \pm .0057$). The highest $F_{ST}$ distances for the Yemeni population were observed with southern and southeastern Africans (fig. 3; table A7 [online only]). Consistent with that, the admixture analysis showed the Yemeni population as a hybrid of predominantly Ethiopian and Near Eastern maternal gene pools, which provides no significant support for gene flow from Mozambique (table 2).

It is possible, however, that the general-population statistic parameters can be misleading in the estimation of realistic contributions of any specific temporal episode of gene flow, if there are multiple migrations. Reconstruction of population histories also becomes complicated when any of the potential source populations has received gene flow from another competing source, or, even more, from a population that is itself considered as the hybrid. Here, both Ethiopian and Yemeni populations can be considered to be hybrids of gene flow from sub-Saharan Africa and the Near East. Therefore, the proximity of Yemenis to Ethiopian and Egyptian populations in the MDS plot and the insignificant southeastern African contribution revealed by the admixture analysis could reflect only the trivial fact that the northeastern African and Yemeni populations are all hybrids of the same basic components. To focus on what could have been the specific elements of recent gene flow across the Red Sea region, we performed phylogeographic founder analysis for the haplotypes observed in Ethiopians and Yemenis, derived from the sub-Saharan African and western Eurasian mtDNA haplogroup components. If we consider a mutation rate of one transition, on average, per 20,180 years within the region 16090–16365 of the HVS-I, it would be expected that <1% of the lineages in the hybrid population that are derived from a source within the last couple of thousand years would have developed two or more mutations. Therefore, to spotlight possible episodes of gene flow in the Red Sea region since the Axumite period, it is reasonable to consider only haplotypes that have matching or one-step different counterparts in the putative source and hybrid populations. Preconditions and caveats of such an approach are discussed in detail in the study by Richards et al. (2000).

### Sub-Saharan African mtDNA Component in Ethiopians and Yemenis

One-third of the Ethiopian and 37% of Yemeni L lineages are captured by diverse subclades of L3. The most frequent among them, haplogroup L3f, which we propose to define by a combination of transitions at nps 16209 (Salas et al. 2002) and 4218 (fig. 2A), was found
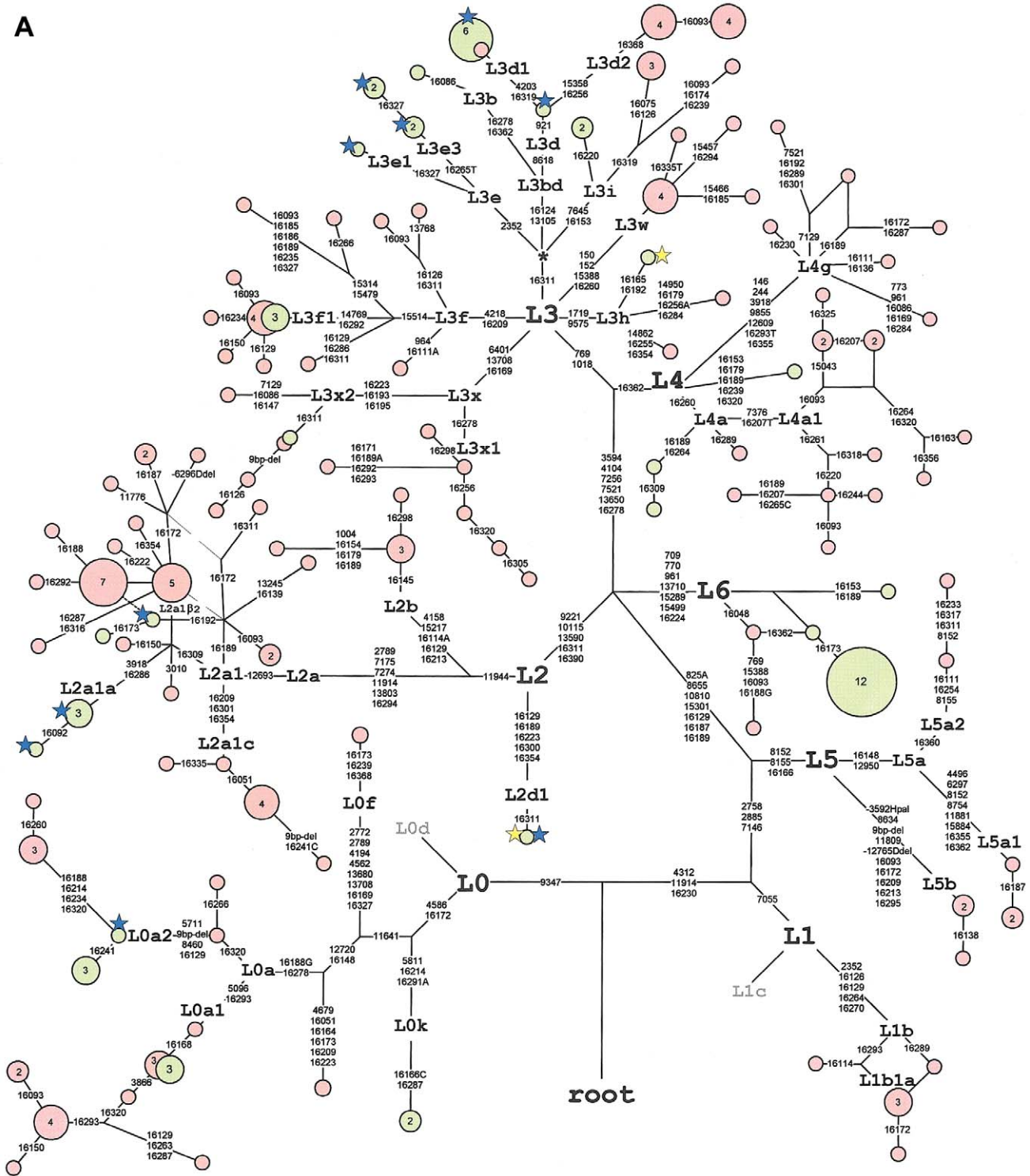
**Figure 2** Median joining network of Ethiopian and Yemeni mtDNA haplotypes. Node sizes are proportional to haplotype frequencies, indicated within nodes for $n > 1$. Haplotypes observed in Ethiopian and Yemeni samples are distinguished by pink and green, respectively. Variable positions (tables A1–A6 [online only]) are indicated along links that connect haplotypes. Nucleotide changes are specified only in the case of transversions. *A,* Network of haplogroups L0–L6. For haplotypes observed in the Yemeni population, matching HVS-I types in samples from northeastern Africa (Krings et al. 1999; Stevanovitch et al. 2004) and Mozambique (Pereira et al. 2001; Salas et al. 2002) are indicated by yellow and blue stars, respectively. *B,* Network of haplogroups M and N. For haplotypes observed in the Ethiopian population, matching HVS-I types in samples from northeastern Africa (Krings et al. 1999; Stevanovitch et al. 2004) and in the Arabian Peninsula and Iraq (Richards et al. 2000; Al-Zahery et al. 2003) are indicated by yellow and blue stars, respectively.
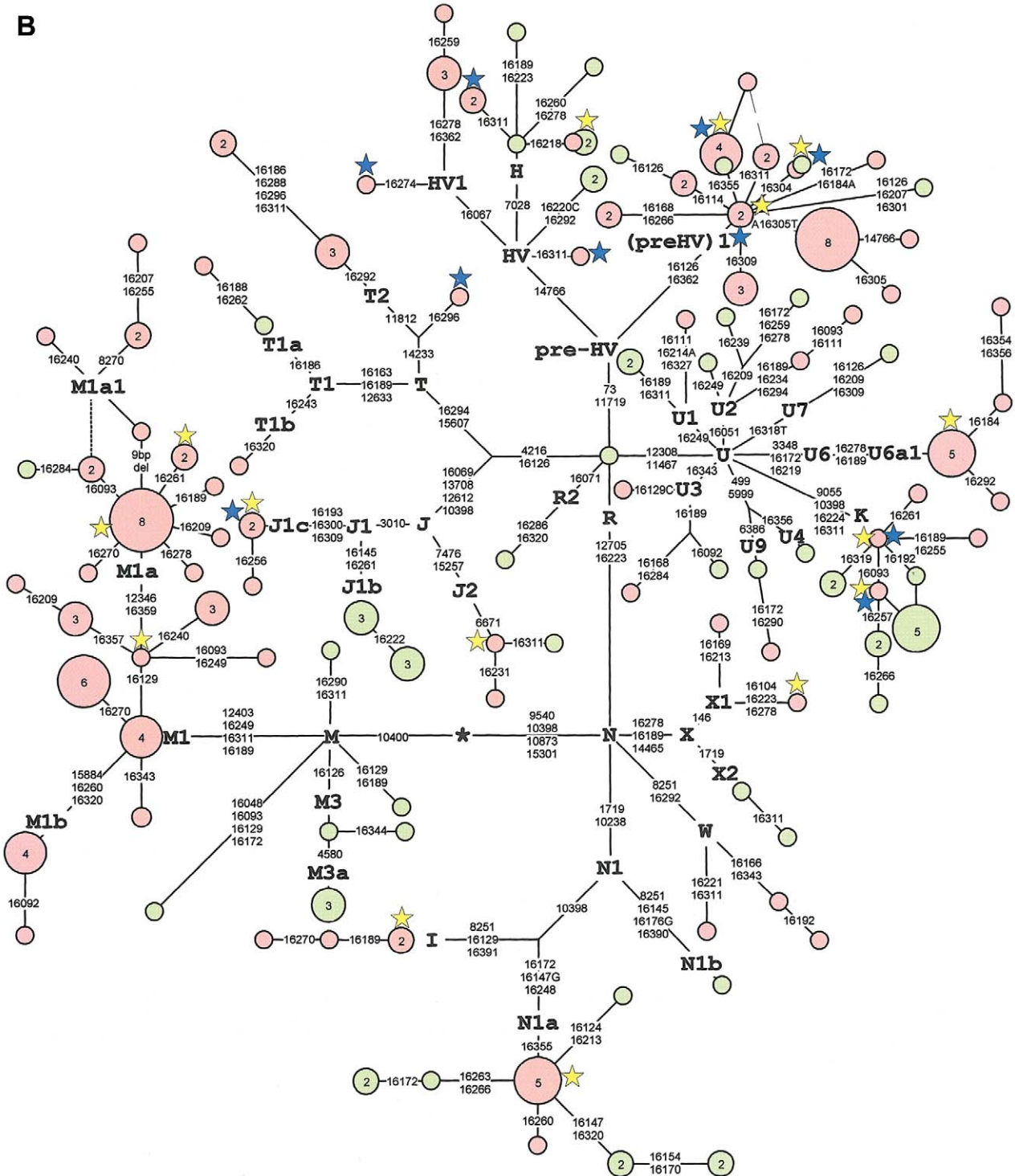
**Table 1**

**mtDNA Haplogroups in Yemen and East Africa**

| HAPLOGROUP | No. (%) of Population | | | | | | |
|---|---|---|---|---|---|---|---|
| | Cushitic Speakers | Amharas | Gurages | Tigrais | Ethiopians[a] | Yemenis | Mozambicans[b] |
| (preHV)1 | 6 (12) | 13 (11) | 2 (10) | 7 (13) | 28 (1) | 4 (3) | |
| H | | 2 (2) | | 1 (2) | 3 (1) | 5 (4) | |
| HV1 | 2 (4) | 1 (1) | | 2 (4) | 5 (2) | | |
| HV* | | | 1 (5) | | 1 (.4) | 2 (2) | |
| J | 1 (2) | 1 (1) | | 3 (6) | 5 (2) | 7 (6) | |
| N1 | | 4 (3) | 2 (10) | 2 (4) | 10 (4) | 9 (8) | |
| T | | 5 (4) | | 3 (6) | 8 (3) | 1 (1) | |
| U6a1 | 1 (2) | 2 (2) | 1 (5) | 2 (4) | 8 (3) | | |
| K | 2 (4) | | 2 (10) | | 4 (1) | 11 (10) | |
| U1–U4, U7–U9 | 1 (2) | 2 (2) | | 3 (6) | 6 (2) | 9 (8) | |
| W | | 1 (1) | | 1 (2) | 3 (1) | | |
| X | | 1 (1) | | 1 (2) | 2 (1) | 2 (2) | |
| R* | | | | | | 2 (2) | |
| N* | | | | | | | 1 (.2) |
| M1 | 7 (14) | 19 (16) | 3 (14) | 9 (17) | 46 (17) | 1 (1) | |
| M* | | | | | | 8 (7) | |
| L0(pre)a | | | | 1 (2) | 1 (.4) | | |
| L0a* | 1 (2) | 1 (1) | | | 2 (1) | | 2 (.5) |
| L0a1 | 2 (4) | 7 (6) | | 2 (4) | 13 (5) | 3 (3) | 40 (10) |
| L0a2 | | 3 (3) | | | 4 (1) | 4 (3) | 61 (15) |
| L0d | | | | | | | 21 (5) |
| L0f | | 1 (1) | | | 1 (.4) | | |
| L0k | | | | | | 2 (2) | |
| L1b | 3 (6) | 1 (1) | 1 (5) | | 6 (2) | | 5 (1) |
| L1c | | | | | | | 22 (5) |
| L2a | 7 (14) | 18 (15) | 2 (10) | 3 (6) | 34 (13) | 6 (5) | 137 (33) |
| L2b | | 4 (3) | | 1 (2) | 5 (2) | | 6 (1) |
| L2c | | | | | | | 3 (1) |
| L2d1 | | | | | | 2 (2) | 3 (1) |
| L3b | | | | | | 1 (1) | 12 (3) |
| L3d1 | 1 (2) | | | | 1 (.4) | 6 (5) | 14 (3) |
| L3d2 | 2 (4) | 3 (3) | | | 8 (3) | | |
| L3d* | | | | | | 1 (1) | 9 (2) |
| L3e | | | | | | 5 (4) | 64 (15) |
| L3f1 | 1 (2) | 3 (3) | 1 (5) | 3 (6) | 8 (3) | 3 (3) | 3 (1) |
| L3f* | 3 (6) | 3 (3) | | | 6 (2) | | 7 (2) |
| L3h | | 2 (2) | | | 2 (1) | 1 (1) | |
| L3i | 1 (2) | 3 (3) | | | 4 (1) | 2 (2) | |
| L3w | | 4 (3) | 1 (5) | 2 (4) | 7 (3) | | |
| L3x1 | 3 (6) | 1 (1) | | 1 (2) | 6 (2) | | |
| L3x2 | 1 (2) | 2 (2) | 1 (5) | | 4 (1) | 1 (1) | |
| L4* | | | | | | 1 (1) | 4 (1) |
| L4a | 2 (4) | 7 (6) | | 4 (8) | 13 (5) | 2 (2) | |
| L4g | | 1 (1) | 1 (5) | 2 (4) | 6 (2) | | |
| L5 | 2 (4) | 4 (3) | 2 (10) | | 8 (3) | | 2 (.5) |
| L6 | | 1 (1) | 1 (5) | | 2 (1) | 14 (12) | |
| Total | 49 | 120 | 21 | 53 | 270 | 115 | 416 |

NOTE.—Haplogroup definitions are according to figure 2 and tables A1–A6 (online only).

[a] Includes all Ethiopians from this study.

[b] Data are from Pereira et al. (2001) and Salas et al. (2002).

in 14 Ethiopian and 3 Yemeni samples. Monophyletic descent of its major subset L3f1, previously characterized by its HVS-I motif alone (Salas et al. 2002), is now supported by two coding-region mutations (fig. 2*A*). With only a few matches to its founder lineages in Central and West Africa (Salas et al. 2002; Rosa et al. 2004), the spread zone of haplogroup L3f appears to be mostly in East Africa. Three Yemeni samples share the founder haplotype of L3f1 with four Ethiopian samples (three Amharas and one Tigrai) and do not show any subsequent differentiation. Although the presence of diverse haplotypes of this clade in West Africa points to an early

rather than a recent dispersal of L3f1 lineages from the east, the founder haplotype of L3f1 has not been sampled among Bantu-speaking populations of southeastern Africa (Salas et al. 2002).

On the basis of the information from coding- and control-region sequencing, we define and characterize three novel L3 subclades—L3i, L3x, and L3w. Haplogroup L3i is characterized by a transition at np 7645 and an HVS-I motif 16153–16223 that was observed in thee Amhara, one Oromo, and two Yemeni samples. It is intriguing that this HVS-I motif, in combination with an additional mutation at np 16319 that was found in all our L3i Ethiopian samples, occurs in parallel in a sister group of haplogroup W lineages in Eurasia (Derbeneva et al. 2002). This highlights the need for assessing coding-region markers, to be able to define mtDNA haplogroups without ambiguity. Three Sudanese HVS-I sequences (Krings et al. 1999), with the motif 16129-16153-16223, are potentially related to L3i. On the other hand, none of the available 516 southeastern African HVS-I sequences (Pereira et al. 2001; Salas et al. 2002; Knight et al. 2003) are likely to belong to this haplogroup.

Haplogroup L3x is defined by transitions at nps 6401, 13708, and 16169 and was found in 10 Ethiopians, with most frequent appearance (12%) among the Oromos. Two subclades of L3x can be distinguished by HVS-I motifs (fig. 2A). The divergent L3x1 and L3x2 subclades seem to be restricted to the Horn of Africa and the Nile Valley, since they have not been observed among southeastern, West, or Central African populations (Watson et al. 1997; Rando et al. 1998; Pereira et al. 2001; Salas

**Table 2**

Estimated Admixture Proportions of Yemeni mtDNA Lineages

| Parental Population | Percentage (SD) of mY |
|---|---|
| Ethiopia ($n = 270$) | 62 (12) |
| Mozambique[a] ($n = 416$) | 2 (8) |
| Near East[b] ($n = 298$) | 36 (8) |

NOTE.—Admixture coefficients (mY) were calculated in the ADMIX 2.0 program from haplogroup frequencies without taking into account molecular distances between haplogroups (Dupanloup and Bertorelle 2001). Bootstrap estimates of the coefficients and their SEs were calculated using 1,000 bootstrap replicates. Haplogroup structure used is as defined in table 1.

[a] Data are from Pereira et al. (2001) and Salas et al. (2002).

[b] Data are from Richards et al. (2000) (including Iraq, Palestine, and Syria).

et al. 2002; Knight et al. 2003; Stevanovitch et al. 2004). An exact match with an Amharan sequence was observed in a Yemeni sample in the L3x2 cluster.

Haplogroup L3w occurred in seven Ethiopian samples but was not found in Yemenis. This clade is defined by substitutions at nps 15388 and 16260. Notice that the coding-region information is obligatory for distinguishing L3w from L4a, where the 16223-16260 HVS-I motif is recurrent (fig. 2A). On the basis of available HVS-I information, both L3w and L4a seem to be restricted to East and northeastern Africa, where they can be detected at low frequencies (Watson et al. 1997; Krings et al. 1999).
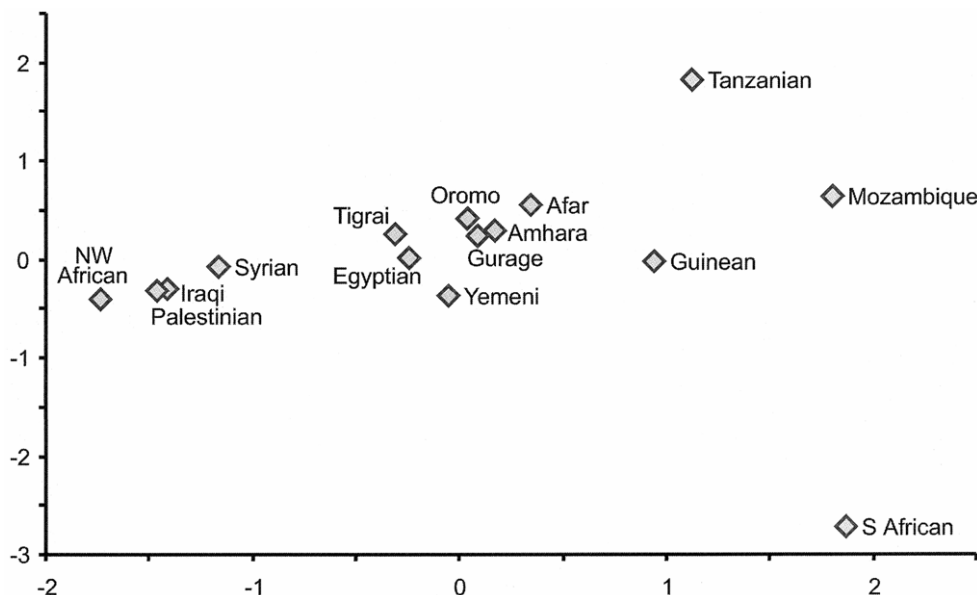


**Figure 3** MDS plot of population differences. The plot is based on $F_{ST}$ distances (table A7 [online only]) calculated from haplogroup frequencies; its stress value is 0.065.

**Table 3**

**Coalescent Times and Haplotype Diversities in mtDNA Haplogroups Observed in Ethiopia and Yemen**

| | ANALYSIS OF MTDNA HAPLOGROUPS IN | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Ethiopia | | | | Yemen | | | |
| HAPLOGROUP | $n$ | $H^a$ (SE) | $\rho$ (SE) | $T^b$ (SE) | $n$ | $H^a$ (SE) | $\rho$ (SE) | $T^b$ (SE) |
| (preHV)1 | 28 | .87 (.05) | .71 (.26) | 14.4 (5.3) | | | | |
| U6a1 | 8 | .64 (.18) | .63 (.33) | 12.6 (6.7) | | | | |
| K | | | | | 11 | .78 (.11) | 1.82 (.83) | 36.7 (16.8) |
| M1 | 46 | .93 (.02) | 2.04 (.84) | 41.2 (17) | | | | |
| M* | | | | | 8 | .89 (.11) | 1.63 (.72) | 32.8 (14.5) |
| L0a1 | 13 | .87 (.07) | 2.54 (1.25) | 51.2 (25.1) | | | | |
| L1b | 6 | .8 (.17) | 2 (1.11) | 40.4 (22.3) | | | | |
| L2a | 34 | .93 (.03) | 2.82 (1) | 57 (20.3) | 6 | .8 (.17) | 2.33 (1.08) | 47.1 (21.8) |
| L3d2 | 8 | .57 (.09) | .5 (.5) | 10.1 (10.1) | | | | |
| L3e | | | | | 5 | .8 (.16) | .6 (.45) | 12.1 (9) |
| L3f1 | 8 | .79 (.15) | .5 (.25) | 10.1 (5) | | | | |
| L3f* | 6 | 1 (.1) | 2.5 (.73) | 50.5 (14.7) | | | | |
| L3w | 7 | .71 (.18) | .29 (.2) | 5.8 (4.1) | | | | |
| L3x | 10 | 1 (.04) | 3.4 (1.06) | 68.6 (21.5) | | | | |
| L4a | 13 | .97 (.04) | 2.54 (.85) | 51.2 (17.2) | | | | |
| L4g | 6 | 1 (.1) | 2 (.62) | 40.4 (12.6) | | | | |
| L5 | 8 | .93 (.08) | 4.75 (1.3) | 95.9 (26.2) | | | | |
| L6 | | | | | 14 | .27 (.15) | 1.93 (1.27) | 38.9 (25.6) |

NOTE.—Estimates are provided only for haplogroups that include five or more individuals.

[a] Nei's haplotype diversity.

[b] Coalescent times ($T$ [ × 1,000]) and their SEs are calculated by $\rho$, according to guidelines of Forster et al. (1996) and Saillard et al. (2000) and under the assumption of a mutation rate of one transition per 20,180 years, in the range of nps 16090–16365 in HVS-I.

Haplogroup L3h lacks a distinctive HVS-I motif and is defined by substitutions at nps 1719 and 9575. One of its subsets with HVS-I motif 16129-16223-16256A-16311-16362 was first reported at a moderate frequency in Guinea-Bissau populations of West Africa (Rosa et al. 2004). In the present study, two different L3h haplotypes were found in Amharic speakers, one of which shares the 16256 transversion with West Africans but lacks the 16129 and 16362 transitions (fig. 2*A*). A related sequence that shares combined transitions at nps 16179 and 16284 has been sampled in Tanzania (Knight et al. 2003). The Yemeni L3h haplotype shares its characteristic HVS-I motif 16165-16192-16223-16311 with six Sudanese sequences (Krings et al. 1999).

Given the geographic distribution of L3f, L3h, L3i, and L3x lineages in East Africa, their presence in Yemen is more consistent with a recent gene flow from Ethiopians or Nile Valley populations than from southeastern Africa.

No L3b or L3e lineages that are widely spread in West African populations and Bantu-speaking southeastern Africans (Salas et al. 2002; Rosa et al. 2004) were found in Ethiopians. In contrast, five Yemeni sequences belonged to haplogroup L3e, consistent with gene flow mediated by the Arab slave trade from southeastern Africa, where the exact matches to Yemeni L3e haplotypes can be found (fig. 2*A*). Haplogroup L3d, with a predominantly West

African distribution (Salas et al. 2002), was present in nine southern Ethiopians. One Oromo among them had a haplotype match with six Yemeni sequences in the L3d1 subclade. This has a characteristic 16124-16223-16319 HVS-1 motif, which shows an elevated frequency in southeastern Africa. This haplotype coexists in a Bantu-speaking Sukuma sample from Tanzania (Knight et al. 2003) with L0a2 lineages characteristic of many Bantu-speaking populations. Another eight L3d lineages in Ethiopians cluster in the L3d2 clade, defined by transitions at nps 15358 and 16256 (fig. 2*A*). The ancestral HVS-I motif of L3d2 has been observed in West African populations (Graven et al. 1995; Watson et al. 1997; Rosa et al. 2004). All eight Ethiopian L3d2 lineages, however, with low haplotype and sequence diversity (table 3), can be distinguished from their West African relatives by a transition at np 16368 (fig. 2*A*) that has not been found, so far, elsewhere in Africa.

The derived allele at np 3594 was previously taken as the defining marker for haplogroup L3 (Watson et al. 1997). By use of information from the coding region, it is now possible to distinguish an early branch from L3, L4, whose descendants share the ancestral states at two nucleotide positions—769 and 1018—and are derived at np 16362 (fig. 2*A*). A combination of three coding- and three control-region markers splits haplogroup L4 into two major subclades. Haplogroup L4a1 is defined

by substitutions at nps 195, 198, 7376, 16207, and 16260 and occurs in 12 Ethiopians with high haplotype and sequence diversity (table 3). Previously, an HVS-I sequence with the L4a1 motif has been observed only in Sudan and Ethiopia (Krings et al. 1999; Thomas et al. 2002), although it was likely misclassified as a member of haplogroup L3e4 (Salas et al. 2002). It is notable that two Yemeni samples and one Ethiopian L4a sample lack the characteristic coding-region marker of L4a1 at np 7376 and also have the ancestral state at HVS-I position 16207. The particular combination of HVS-I mutations of 16223-16293T-16311-16355-16362, recently described by Salas et al. (2002) as a haplogroup L3g motif, shows an almost exclusively East African distribution and a coalescent time of 40,400 years ± an SE of 12,600 in Ethiopians (table 3). This motif, together with two characteristic mutations at nps 146 and 244 from HVS-II, has been found at its highest frequency (52%) in the Hadzabe population from Tanzania, but without any haplotype diversity ($h = 0$; $n = 19$) (Vigilant et al. 1991; Knight et al. 2003). In contrast, highly diverse Ethiopian sequences (table 3) with the combined L3g HVS-I and HVS-II motifs share the ancestral character states, at nps 769 and 1018, with haplogroup L4a and are derived at five other positions that distinguish L3 from L2 (table 1; fig. 2A). This information—plus the sharing of the 16362C allelic state—suggests that L3g is actually a sister clade of L4a; therefore, we propose to rename it "L4g" (fig. 2A). Only a single Yemeni L4 lineage, which shares no common mutations with Ethiopian L4 mtDNAs, was found to belong to haplogroup L4.

Both L4a and L4g (previously named "L3g") reveal high haplotype and sequence diversity in Ethiopians. The coalescent calculations suggest that L4 lineages diversified from their founder 68,800 ± 18,300 years ago. That date is within the error margin of the HVS-I (77,000 ± 2,400 years) dating of the whole L3 cluster (Watson et al. 1997). Taken together, these findings suggest that the origin and split between L4 and L3 lineages occurred during the late Pleistocene in East Africa, in the Red Sea region, in a period that is close to the proposed beginning of the colonization(s) of Eurasia and the rest of the world—~50,000–60,000 years ago—by the ancestors of the two extant daughter groups of L3 (Stringer 2000; Forster 2004).

All Ethiopian L2 lineages can be seen as derived from the two subclades L2a1 and L2b (fig. 2A). L2a1, defined by the mutation at np 12693, includes two further minor subclades: L2a1a, as defined by a substitution at np 16286 (Salas et al. 2002), is now supported by a coding-region marker (np 3918) (fig. 2A) and was found in four of six Yemeni L2a1 lineages. L2a1a occurs at its highest frequency in southeastern Africa (Pereira et al. 2001; Salas et al. 2002). Both the frequent founder haplotype and derived lineages (with 16092 mutation) found among Ye-

menis have exact matches within Mozambique sequences (Pereira et al. 2001; Salas et al. 2002). Most Ethiopian L2a1 sequences share mutations at nps 16189 and 16309 (L2aβ2 [Salas et al. 2002]), and a minor portion, L2a1c, shares mutations at nps 16209, 16301, and 16354 (within cluster L2a α1 [Salas et al. 2002]). The L2a1β2 HVS-I motif shows a pan-African spread (Salas et al. 2002). However, whereas the majority (26/33) of African American L2a complete sequences could be partitioned into four subclades by substitutions at nps 3495, 3918, 5581, and 15229 (Torroni et al. 2001; Howell et al. 2004), none of these were observed in our Ethiopian L2a1 samples. A single L2d1 sequence from the Yemeni sample shares the haplotype that has so far been observed in Sudan and in southeastern Africa (Salas et al. 2002). Consistent with figure 7a of Salas et al. (2002), Ethiopian L2b sequences form a subset of a predominantly West African clade, distinguished from West African lineages by a transition at np 16145.

A novel sister clade of L2 and L3′4, defined by six coding transitions and one control-region transition, was found in two Ethiopian and 14 Yemeni (12%) samples. We propose to label it "L6" (fig. 2A). It is interesting that the highly frequent L6 haplotype observed in the Yemeni sample does not have a close match among >3,000 African HVS-I sequences (Salas et al. 2002; Rosa et al. 2004). An East African origin of haplogroup L6 seems most likely, because of its presence in Ethiopians and the fact that its sister haplogroups L2, L3, and L4 are all diverse and frequent there. Given the lack of an exact match from the African database for the southern Arabian L6 samples and the relatively deep time depth of its variation in Ethiopians and Yemenis—taken together, 36,600 ± 23,400 years—it is possible that this haplogroup has been preserved in isolation in the Ethiopian Highlands and southern Arabia for tens of thousands of years. However, the most frequent haplotype of L6 in Yemenis does not bear any descendant lineages, which suggests that its carriers coalesce to a common ancestor within only a couple of thousands of years.

Haplogroup L1, which is more frequent and diverse in West and Central Africa than in East Africa (Salas et al. 2002), is represented in Ethiopia by six L1b lineages, whereas L1c is completely absent in our Ethiopian and Yemeni samples. Five Ethiopian L1b lineages share a transition at np 16289—which defines a founder, L1b1a—that is spread among Ethiopians and Nubians (Krings et al. 1999) and is associated with relatively low downstream variation (fig. 2A).

A further cluster, L1e—previously characterized in the HVS-I–based classification of Salas et al. (2002) and found among the Gurnas of Egypt (Stevanovitch et al. 2004) and the Sukuma of Tanzania (Knight et al. 2003)—has been recently redefined as "L5a" because it occupies an intermediate phylogenetic position between L1 and

the L2′L3 clades (Shen et al. 2004). The spread of this haplogroup is restricted to East Africa (Salas et al. 2002). Five L5a and three L5b lineages can be distinguished in our Ethiopian sample (fig. 2). The geographic spread of the L5b lineages is more southern, extending to the Sukuma from Tanzania (Knight et al. 2003).

Haplogroup L0 is the earliest offshoot of the mtDNA tree in Africa that appears as a sister group to the branch that holds all other haplogroups found in extant humans (fig. 2A). It includes four subclades (L0a, L0d, L0f, and L0k) that were previously classified as subclades of the paraclade L1 (Watson et al. 1997; Salas et al. 2002). L0 is represented in the Ethiopian sample primarily by its daughter clade L0a and a single mtDNA genome from haplogroup L0f. L0a1 accounts for the majority of L0 lineages in Ethiopians, whereas L0a2, widespread in Africa (Salas et al. 2002), is associated with the 9-bp deletion in the COII/tRNA$^{Lys}$ intergenic region and appeared to be the predominant L0a subclade in the Yemenis. The four L0a2 HVS-I sequences in Ethiopians differed by four substitutions from the frequent modal haplotype of the Bantu speakers (Soodyall et al. 1996; Salas et al. 2002). This haplotype is also present in our Yemeni sample. It is notable that the 9-bp deletion was observed, in parallel, against three other phylogenetic backgrounds of Ethiopian mtDNA sequences in haplogroups M1, L2a1, and L3x.

Three Ethiopian samples showed neither L0a1- nor L0a2-defining mutations and thus remain unclassified at the L0a level. One of them even lacked the 16188 transversion characteristic of L0a but shared both its defining coding-region mutation at np 12720 and the 16148 mutation in HVS-I. Control-region sequences of the L0f type have been found so far at marginal frequencies only in East Africa, with the highest incidence (3/12) among the Iraqw population of Tanzania (Watson et al. 1997; Knight et al. 2003). The phylogeny of the L0 clade in Ethiopians lends further credence to the idea that East Africa is the most likely source of haplogroup L0a variation (Salas et al. 2002). Besides L0a and L0f, this most ancient trunk of the human mtDNA tree includes two subclades—L0d and L0k—that are specific to Khoisanspeaking populations in South Africa (Chen et al. 2000). It is interesting that two L0k sequences were recovered from the Yemeni sample, which indicates gene flow from southeastern Africa.

Taken together, lineages of haplogroups L0–L5 and their subclades reveal high regional specificity in Africa, with haplotypes observed in Ethiopians being, by and large, different from those spread in southeastern Africa—in particular, those in Mozambique. The pattern of shared deeply rooting lineages that are distant in their inner branches is analogous to that observed for Y-chromosomal phylogenies (Semino et al. 2002). This finding is supported by the significant $F_{ST}$ differences observed

between those two regions of Africa. However, Yemeni sequences, in contrast to their clustering together with Ethiopian and Egyptian populations in the MDS plot (fig. 3), show greater affiliation to haplotypes detected in Mozambique. The haplotype sharing with Mozambique accounts for 23% of the total and 49% of haplogroup L0–L5 lineages among Yemenis. The lack of M and N lineages in the Mozambique sample is the only apparent factor that separates it from Yemenis in the MDS plot. It should be noted here that the percentage of shared lineages between Yemeni and Mozambique mtDNAs cannot be taken as a measure of actual admixture proportion, because there is a substantial fraction of uninformative haplotypes in both samples. These include either matches or the lack of matches, both in northeastern and southeastern African populations, that probably reflect the incomplete sampling of Africa. Compared with Bantu speakers from southeastern Africa, the Ethiopian contribution to the Yemeni mtDNA pool can be considered relatively minor, since the shared haplotypes account for just 9% of the total variation.

*Haplogroup M Lineages in Ethiopians and Yemenis*

Haplogroup M1 lineages constitute 17% of the Ethiopian mtDNA sequences, consistent with their high frequency in the region (Passarino et al. 1998; Quintana-Murci et al. 1999; Richards et al. 2003). Two subclades, which can be distinguished by coding-region RFLPs (Quintana-Murci et al. 1999)—M1a by 12345 *Rsa*I (12346T) and M1b by 15883 *Ava*II (15884A)—together account for 56% of its variation. M1a is further characterized by a transition at np 16359 in HVS-I and is also present in the single Yemeni M1 sample (fig. 2B). M1a can be found together with M1* lineages in populations from the Near East, the Caucasus, and in Europe at marginally low frequencies (Corte-Real et al. 1996; Macaulay et al. 1999; Richards et al. 2000). The minor group M1b, defined by the motif 15884-16260-16320, is restricted to East Africans, having been observed, so far, only in Ethiopians (Quintana-Murci et al. 1999) and in Egypt (authors' unpublished results). It is interesting that the variable noncoding nucleotide 15884 also carries the derived A allele in one Moroccan M1* complete sequence, yet without the characteristic M1b HVS-I pattern (Maca-Meyer et al. 2001). M1a and M1b sequences are rare or absent in North Africans (Corte-Real et al. 1996; Rando et al. 1998; Brakez et al. 2001; Plaza et al. 2003). Instead, a third clade, M1c, defined by a transition at np 16185, covers most of haplogroup M1 variation in northwestern Africa, the Canary Islands, and the Near East. M1c has not been sampled yet among Ethiopians. It is intriguing that a Moroccan M1c complete sequence (Maca-Meyer et al. 2001) lacks the 813-6671-12950C mutations that define a common branch holding

the M1a and M1b clades (fig. 4). It is notable that the other Moroccan M1 sequence with the 15884 mutation also lacks the 6671-12950C signature. In light of these data and because of the lack of other distinctive East African–specific mtDNA haplogroups in northwestern Africa, it is difficult to interpret the northwestern African haplogroup M1 variation as a derivative from the East African mtDNA pool.

Yemeni M sequences show matches with some Indian sequences—for example, in M3 (Kivisild et al. 1999, 2003*a*; Bamshad et al. 2001). Their presence probably reflects recent gene flow, consistent with the historical fact that southern Yemen was under the rule of British India during 1839–1937 and that a substantial population of South Asians can be found in southern Yemen today.

### Haplogroup N Lineages in Ethiopians and Yemenis

Lineages that belong to haplogroup N that cover virtually all mtDNA sequences in western Eurasia (Richards et al. 2000) show substantial frequencies both in the Yemeni (44%) and Ethiopian (31%) mtDNA pools. In this respect, Ethiopians differ explicitly from most other sub-Saharan African populations studied thus far. Within Ethiopia, the frequency of N lineages is significantly higher ($P < .05$) in samples that originate from its northern territory (48%), which was the center of the Aksum kingdom, than among other Ethiopians, mostly originating from the south-central part of the country (27%). At the same time, there was no significant difference in the proportions of haplogroup N between the Semitic and Cushitic linguistic groups in our sample—for example, between Amharas and Oromos.

Haplogroup (preHV)1 is by far the most frequent (10.4%) subclade in the Ethiopian N cluster (fig. 2*B*). The majority of the Ethiopian (preHV)1 lineages match or derive from founder haplotypes common to Near Eastern, southern Caucasian, and North African populations (Krings et al. 1999; Metspalu et al. 1999; Richards et al. 2000; Kivisild et al. 2003*b*). Previously, the highest frequency (20.4%) of (preHV)1 lineages was observed in Yemeni Jews (Richards et al. 2003), significantly higher than their frequency in our Yemeni non-Jewish sample (3.4%; $P < .01$). This probably reflects strong genetic drift in the founding population of Yemeni Jews. Because (preHV)1 lineages occur in populations of the Near East, the Caucasus, and Mediterranean Europe—where African L0-L6 lineages are absent or rare—it is more likely that their presence in East Africa reflects a back-migration from the Near East rather than an in situ origin of (preHV)1 in Ethiopia (Richards et al. 2003). Nevertheless, we notice that several Ethiopian (preHV)1 lineages, including (1) variants with a transversion at np 16305, (2) HVS-I motif 16126-16309-16362, and (3) HVS-I
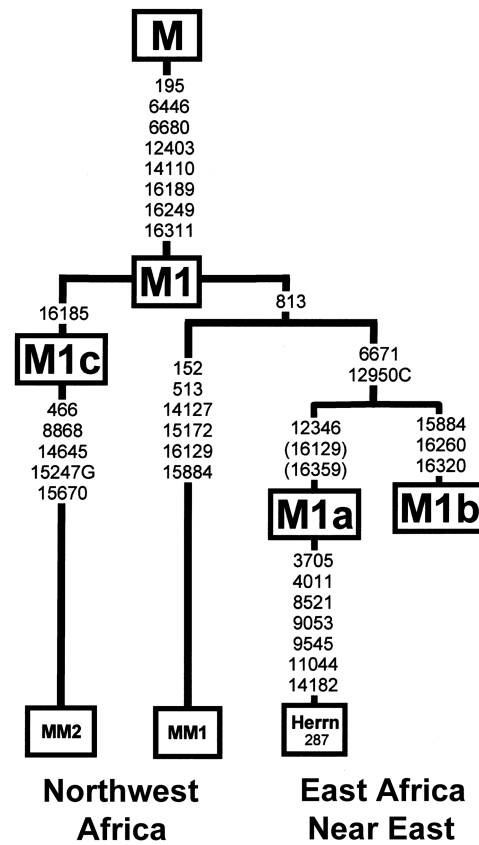


**Figure 4** Phylogeny of mtDNA complete sequences of haplogroup M1. The tree and the character changes are reconstructed by use of maximum-parsimony criteria. Mutations are numbered relative to revised Cambridge Reference Sequence (rCRS) (Andrews 1999). Nucleotide change is specified only for transversions. Sources: MM1 and MM2 = Maca-Meyer et al. (2001); Herrn 287 = Herrnstadt et al. (2002). Haplogroups M1a and M1b are defined as in figure 2*B* and tables A1–A6 (online only), by polymorphisms identified by Quintana-Murci (1999). Character-state changes at nps 16129 and 16359, shown in parentheses on the external branch of the Herrnstadt et al. (2002) coding-region sequence, are inferred from the complete linkage of the 12346 polymorphism with the derived status at these positions in the present study.

motif 16126-16172-16184A-16362, were not found in 185 (preHV)1 sequences sampled from >20,000 individuals from Arabia, the Near East, and Europe (Macaulay et al. 1999; Metspalu et al. 1999; Richards et al. 2000; authors' unpublished data), except for an HVS-I haplotype 16126-16305T-16362 that occurs (12.5%) in Ethiopian Jews (Thomas et al. 2002). Their elevated frequency and uniform presence among major language groups in Ethiopia (table 1) suggests that these derived lineages may represent a relatively old introgression of lineages to the Ethiopian mtDNA pool from the Near East.

Haplogroup HV1 is represented in Ethiopians by two different HVS-I motifs (fig. 2*B*). The first of them, 16067-

16274, observed in an Amharan mtDNA, has been reported in populations from the Arabian Peninsula (Di Rienzo and Wilson 1991; Richards et al. 2000), southern Egypt, and northern Sudan (Krings et al. 1999). The other four sequences, present in Tigrais and Oromos, share a common HVS-I motif, 16067-16278-16362, that has not yet been reported in the literature. Two Yemeni HV* samples belong to a cluster of sequences with the characteristic 16220C transversion, observed more frequently in the Caucasus and the Near East (Richards et al. 2000).

When the fact that haplogroup H is the predominant subclade of N in most western Eurasian populations is considered, its frequency in Ethiopians is surprisingly low (0.7%). Among the three haplogroup H lineages found, one Tigrai carried a characteristic HVS-I transition at np 16218, which has been observed in haplogroup H lineages—mostly in those of Near Eastern origin, but also in two Yemeni H sequences and two Assiut sequences from Egypt (Krings et al. 1999; Richards et al. 2000).

Three of the five haplogroup J lineages in Ethiopians share a distinct HVS-I motif, 16069-16126-16193-16300-16309 (J1c), that is characteristic of J sequences in populations from the southern Caucasus, the Near East, and North Africa (Di Rienzo and Wilson 1991; Richards et al. 2000; Brakez et al. 2001; Maca-Meyer et al. 2001; Plaza et al. 2003). In East Africa, J1c sequences have been found in one Datoga from Tanzania (Knight et al. 2003) and in one Gurna from Egypt (Stevanovitch et al. 2004). The other two Ethiopian J sequences, present in Tigrais, belonged to a subclade of J2 that is defined by a transition at np 6671 (Herrnstadt et al. 2002). Most of the Yemeni J sequences, in contrast, share the combination of 16145 and 16261 mutations in haplogroup J1b, which is a common motif of J lineages in populations from the Near East and all over western Eurasia (Richards et al. 2000).

All Ethiopian and Yemeni haplogroup T sequences clustered with either T1 or T2 subclades, consistent with the classification of all existing European T coding-region sequences (Ingman et al. 2000; McMahon et al. 2000; Finnilä et al. 2001; Herrnstadt et al. 2002; Coble et al. 2004). One Amhara T sequence, however, which harbors a transition at np 14233, characteristic of T2 sequences, lacked the other substitution at np 11812, present in all other Ethiopian and European T2 sequences. The np 11812 substitution was similarly absent in a complete North African T sequence (Maca-Meyer et al. 2001). The Tigrai T1a sequence matches a Kerma sequence from Nubia (Krings et al. 1999), whereas the Amhara T1b sequence shows a mutation at np 16320 on top of the common founder haplotype in the Near East (Richards et al. 2000). Five of the six T2 sequences detected among Amhara and Tigrai samples shared a transition at np 16292 that is widespread in the haplo-

group T context in Europe, the Near East, and North Africa. However, the two Tigrai T2 sequences share a combination of four downstream HVS-I mutations (fig. 2B) that have not been reported elsewhere.

N1a is a minor mtDNA haplogroup that has been observed at marginal frequencies in European, Near Eastern, and Indian populations (Mountain et al. 1995; Richards et al. 2000). It occurs at a significant frequency in both Ethiopian and Yemeni populations. Six Ethiopian N1a lineages, restricted to Semitic-speaking subpopulations, show low haplotype diversity and include an exact HVS-I sequence match with a published N1a sequence from Egypt (Krings et al. 1999). A related sequence, from southern Sudan (Krings et al. 1999), was misclassified as a member of the L1a clade (Salas et al. 2002). Yemeni N1a sequences, on the other hand, display a high level of haplotype ($h = 0.89$) and nucleotide ($\rho = 2.75 \pm 1$) diversity, combined with the highest frequency (6.9%) of this haplogroup reported so far.

Two Ethiopian haplogroup X sequences from this study have been characterized elsewhere as belonging to North and East African–specific subclade X1 (Reidla et al. 2003). A control-region sequence similar to the Tigrai X1 haplotype was found recently in a Gurna sample from Egypt, though it was probably mislabeled as "L3" by the authors, since no coding-region markers specific to either haplogroup X or L3 were determined in that study (Stevanovitch et al. 2004). Both Yemeni X sequences, in contrast, belong to the major western Eurasian subclade X2.

Haplogroup-U4 lineages are spread at moderate frequencies all over Europe, western Siberia, and southwestern Asia and coalesce to their most recent common ancestor within the Middle or Late Upper Palaeolithic period (Richards et al. 2000; Tambets et al. 2003). On the basis of complete sequence data (Finnilä et al. 2001; Herrnstadt et al. 2002), the consensus sequence of U4 in Europe differs from the root haplotype of U by nine substitutions. One Tigrai and one Yemeni U sample shared a transition at np 6386 that defines U9, which was recently detected in South Asia (Quintana-Murci et al. 2004). Therefore, haplogroups U4 and U9 are sister clades within a clade that is defined by one control- and one coding-region mutation (fig. 2B).

All eight Ethiopian U6 samples descend from the major U6a1 founder (fig. 2B), which is spread from the Near East to northwestern Africa at appreciable frequencies (Maca-Meyer et al. 2003). Their absence in Yemen suggests that these U6 lineages have likely penetrated to Ethiopia from the north rather than by the sea route from Arabia. Conversely, both Amhara and Tigrai U2 sequences share an HVS-I motif, 16051-16189-16234-16294, that has not been sampled, to date, in North Africa but that can be found at low frequencies in popu-

lations of western Asia and the Caucasus (authors' unpublished data).

Despite the fact that haplogroup-N lineages occur at equally high frequency in Ethiopians and Yemenis, only three haplotypes (representing 3% of the total Yemeni sample) were found to match between the two populations. This suggests that the immediate source populations for these lineages in the Near East, from which they derive, could have been different.

## Discussion

### Western Eurasian Genetic Component in Southern Arabia and East Africa

Though present-day Ethiopia is a land of great ethnic diversity, the majority of Ethiopians speak different Semitic, Cushitic, and Omotic languages that belong to the Afro-Asiatic linguistic phylum. Maternal lineages of Semitic- (Amharic, Tigrinya, and Gurage) and Cushitic- (Oromo and Afar) speaking populations studied here reveal that their mtDNA pool is a nearly equal composite of sub-Saharan and western Eurasian lineages. This finding, consistent with classic genetic-marker studies (Cavalli-Sforza 1997) and previous mtDNA results, is also in agreement with a similarly high proportion of western Asian Y chromosomes in Ethiopians (Passarino et al. 1998; Semino et al. 2002), which supports the view (Richards et al. 2003) that the observed admixture between sub-Saharan African and, most probably, western Asian ancestors of the Ethiopian populations applies to their gene pool in general. On the other hand, significant differences in the proportions of derived lineages of haplogroup N between northeastern and south-central samples from Ethiopia are consistent with the proximity of the Tigrinya region (Aksum) and Eritrea to the coast of the Red Sea, the latter having mediated gene flow with Egypt and southern Arabia—perhaps, in particular, with the rise of Semitic cultural influence in the region. In contrast, the similarity of Amharas and Oromos, also expressed in other genetic loci (Fort et al. 1998; Corbo et al. 1999), supports the idea that "amharization" may have been largely a sociocultural rather than a genetic phenomenon. Yet, it is important to add here that Y-chromosomal haplogroup J1-M267, which is widespread throughout Arab-speaking countries and encompasses a third of Amharan Y chromosomes, has hardly penetrated the Cushitic-speaking Oromo population (Semino et al. 2004).

Haplogroups M1 and (preHV)1 occurred at almost equal frequencies among Cushitic- and Semitic-speaking populations of Ethiopia (table 1). Both haplogroups are also common in western Asian Semitic-speaking populations and have occasionally been found in North and northwestern African Berbers (Rando et al. 1998; Rich-

ards et al. 2000, 2003; Plaza et al. 2003), which suggests a correlation with the spread of Afro-Asiatic languages (Forster 2004). High diversity of M1 among Cushitic populations of East Africa and the absence of specific subclades present among them, in Tigrais and in all western Asian populations, point to an ancient diversification of M1 in East Africa, consistent with the East African origin of the main subgroups of Afro-Asiatic languages (Ehret 1995). The ancestral status of Moroccan complete sequences at mtDNA coding-region sites that define the major clades present in Ethiopians, however, leaves open the possibility that M1 had originated in North Africa or the Near East instead and was imported to Ethiopia in the remote past, early enough to allow the rise of subclades frequent in and specific to the Horn of Africa. Haplogroups HV, TJ, U, N1, and W combined, on the other hand, were significantly more frequent among Tigrais than among other Ethiopians, which implies that the major part of these lineages may have been imported relatively recently—for example, along with the expansion of Semitic languages in Ethiopia.

Several mtDNA haplogroups—such as (preHV)1, U6, and some subbranches of L3 that Ethiopians share with North African populations—display coalescent times in the early Holocene (table 3) a similar period to that estimated for North and East African Y chromosomes in haplogroup E3b1-M78, which is abundant and may have originated in Ethiopia (Cruciani et al. 2004; Luis et al. 2004). It is interesting that, like E3b1-M78, these mtDNA haplogroups are infrequent or absent in our Yemeni sample (table 1). Note that the identified time window is close to the proposed division of the Semitic and Cushitic branches of Afro-Asiatic languages (Militarev 2003) and corresponds broadly to the beginning of deep environmental changes in the deserts of the Sahara and the Arabian Peninsula, when those regions recovered from their widest span and most extreme aridity during the Last Glacial Maximum period.

On the other hand, similar to mtDNA haplogroup (preHV)1, Y-chromosomal haplogroup J1-M267 can be identified as the sole branch that is highly abundant in the Near and Middle East and in northeastern and East Africa (Underhill et al. 2000; Semino et al. 2002, 2004; Luis et al. 2004). Higher STR diversity of this Y-chromosomal clade among Europeans and Ethiopians, as compared with populations from northeastern Africa and the Middle East, suggests that it may have reached Ethiopia (and Europe) early in the Holocene, whereas its frequent spread in North Africa and the Middle East may have been driven by the expansion of Arabs since the 7th century (Semino et al. 2004). As for the E3b-M35 subclades, the problem is to fit a proposed >30,000-year-old split between the two major sister clades—J1-M267 and J2-M172 (Semino et al. 2004)— into this scenario. Somewhat indirectly (inferred from figure 1 of Luis et

al. 2004), J1-M267 chromosomes appear to be particularly frequent among southern Arabians (38% in Omanis) and well represented in Egypt (20%). Absence of the corresponding STR variation patterns for the Omani and Egyptian samples does not allow, at present, the inference of which, if either of the two, is likely to be closer to the Ethiopian J1-M267 chromosomes. Nevertheless, a clear asymmetry between E3b1-M78 and J1-M267 chromosomes is seen—the former are rare or absent in southern Arabia, whereas the latter are relatively frequent. Hence, Ethiopians may have been recipients of the southern Arabian J1-M267 chromosomes but have not been efficient donors of the E3b1-M78 chromosomes to southern Arabia, although East Africans may have carried the latter to Egypt and, farther, to Europe via the Levantine corridor. Furthermore, as already mentioned above, there is a profound difference in J1-M267 frequencies between the Semitic-speaking Amharas, who probably arrived relatively recently from Arabia, and the Cushitic-speaking Oromos, among whom the frequency of J1-M267 chromosomes does not exceed 3% (Cruciani et al. 2004). Relevant data for other Ethiopian populations and Yemenis are desired for further exploration of this line of arguments.

A majority of mtDNA lineages of Ethiopian Jews (Falasha or Beta Israel) derive from African-specific clades L0–L5 (Shen et al. 2004), including exact matches with Ethiopians sampled in the present study. Consistent with Y-chromosomal findings, this fact points to extensive admixture of Jews with the local population. A specific haplotype match in haplogroup (preHV)1—which is also widely spread in the Near East—between Ethiopian Jews and non-Jews is more problematic, because it is also possible that the non-Jews obtained the lineage from the Jews. This particular (preHV)1 haplotype, with a rare transversion at np 16305, (1) has not been detected, so far, among other Semitic populations of the Near East; whereas, (2) in Ethiopia, it occurs both among Cushitic and Semitic speakers; and, (3) in Ethiopian Jews, there are many sub-Saharan African lineages from haplogroups L0–L3. It is more likely, therefore, that the matching haplotype does not represent the incursion of Jewish maternal lineages into the Ethiopian gene pool but that this haplotype instead substantiates the extent of Ethiopian admixture in the Falasha population. Taken together, the influx of the elements of the Hebraic culture in the first centuries A.D. probably did not have a major impact on the genetic pool of Ethiopians, and the present-day Jews of Ethiopian descent probably assimilated genes from the local non-Jewish populations through conversion of the latter to Judaism. The other two episodes of intrusion of Semitic influence, related to contacts with southern Arabia, are weakly supported by our data. This is because, among the haplogroup N lineages present in high frequency in the Tigrais and other Ethio-

pian ethnic groups, only a few revealed close relationships with equivalent lineages present in southern Arabia.

Haplogroup U9 is a rare clade in mtDNA phylogeny, characterized only recently in a few populations of Pakistan (Quintana-Murci et al. 2004). Its presence in Ethiopia and Yemen, together with some Indian-specific M lineages in the Yemeni sample, points to gene flow along the coast of the Arabian Sea. Haplogroups U9 and U4 share two common mutations at the root of their phylogeny (fig. 2B). It is interesting that, in Pakistan, U9 occurs frequently only among the so-called "negroid Makrani" population. In this particular population, lineages specific to sub-Saharan Africans occur as frequently as 39%, which suggests that U9 lineages in Pakistan may have an African origin (Quintana-Murci et al. 2004). Regardless of which coast of the Arabian Sea may have been the origin of U9, its Ethiopian–southern Arabian–Indus Basin distribution hints that its diversification from U4 may have occurred in regions far away from the current area of the highest diversity and frequency of haplogroup U4—East Europe and western Siberia.

### Sub-Saharan African Genetic Component in Yemenis

Phylogenetic analysis reveals that the origin of sub-Saharan African mtDNA variants in Yemenis is a mosaic of different episodes of gene flow. Three different passages can be outlined. The first is gene flow, likely mediated by the Arab slave trade from southeastern Africa, as evidenced by exact mtDNA haplotype matches. Such matches account for 23% of the total variation in Yemenis and occur in lineages and lineage groups that cannot be found in Ethiopia and northeastern Africa. Many of these can be traced to the Bantu dispersal; they have their origin in West Africa and supply thereby the upper time limit of 3,000–4,000 years for their departure from southeastern Africa toward Arabia. The sub-Saharan African component of Ethiopians has remained untouched by such influences and may therefore be considered most representative of the indigenous gene pool of sub-Saharan East Africa. The relatively high mtDNA contribution from southeastern Africa is in contrast with only a minor proportion of shared Y-chromosomal lineages between these regions (Thomas et al. 2000; Richards et al. 2003; Cruciani et al. 2004).

Second, the relatively minor proportion of lineages that are shared between Yemenis and Ethiopians suggests that putative slave capturing from Ethiopia was of a lesser extent than that from the southeastern coast of Africa. More generally, the low level (9%) of haplotype sharing within haplogroups L0–L6 between Ethiopians and Yemenis also shows that a directional eastward maternal gene flow, even during the peak of the bipartite Aksum kingdom, must have been relatively minor. There is also no support for significant gene flow from Ethiopia

to Arabia, since the Y-chromosome haplogroups E3b-M35* and E3b-V6, which have a quite visible combined frequency (16%–33%) in Ethiopia, are rare or absent in Arabia (Thomas et al. 2000; Semino et al. 2002; Richards et al. 2003; Cruciani et al. 2004; Luis et al. 2004). Similarly, it can be noticed that, whereas the combined frequency of the deepest clades A and B of the Y-chromosome phylogeny has a range of 10%–60% in Ethiopia and Sudan (Underhill et al. 2000; Semino et al. 2002), comparable to the combined frequency of their sub-Saharan mtDNA component, there is a significant excess of the African mtDNA component in another southern Arabian population, the Omanis (37%) (authors' unpublished results), as compared with the virtual absence of the Y-chromosomal counterpart (Luis et al. 2004).

Third, the high frequency of haplogroup L6 in Yemenis points to an enigmatic link between the southwestern Arabian gene pool with that of East Africa. This haplogroup derives from the phylogenetic tree of sub-Saharan African mtDNA haplogroups but shows only marginal incidence in Ethiopians and is completely absent elsewhere in Africa. Its high frequency in Yemen, together with low haplotype diversity, probably reflects the effect of genetic drift in a small founding population. A recent bottleneck of the general Yemeni population seems unlikely because of the high haplotype variation in other haplogroups (table 3). A founder effect from outside is also not supported, because of the lack of a possible source population outside Yemen, in whom the L6 founder haplotype would be present at a significant frequency. From the present evidence, the possibility cannot be eliminated that this haplogroup may even have originated from the same out-of-Africa migration that carried haplogroups M and N and founded the mtDNA diversity of Eurasia, the Americas, and Oceania. Yet, this scenario would imply a total isolation of a southern Arabian population from the others in that region to explain the absence of L6 types in other populations of the Near East, Arabia, and elsewhere in the world. Alternatively, in consideration of the highly heterogeneous haplogroup composition of individual populations from East Africa (e.g., from Tanzania [Knight et al. 2003]) and the almost complete lack of data from some regions (like Somalia and Kenya), it is possible that the source population of Yemeni L6 varieties has not yet been sampled.

In summation, Ethiopian and Yemeni maternal lineages can be seen as composites of both sub-Saharan and western Eurasian mtDNA haplogroups that coexist in almost equal proportions on both sides of the Red Sea. On the surface, it suggests a very extensive bidirectional gene flow between the two areas, readily supported by historic narratives as well as quantitative population statistics. Founder analysis of individual elements of this composition, however, revealed that, during the last several thousands of years, the populations of the Horn of Africa and southern Arabia, though sharing a minor part of their maternally inherited genomes, had received major demic influences from different sources—which they do not necessarily share—including the Near East, India, and northeastern and southeastern Africa. The presence of a frequent founder sequence type of an ancient and as-yet-uncharacterized haplogroup L6 in the Yemeni population, with no haplotype match in the African data base, intriguingly points to a possibly early gene flow across the Red Sea or to a signal of gene flow from an African population that has not yet been sampled.

## Acknowledgments

## Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

ADMIX 2.0, http://web.unife.it/progetti/genetica/Isabelle/admix2_0.html

ARLEQUIN package, http://lgb.unige.ch/arlequin/

EMBL Nucleotide Sequence Database, http://www.ebi.ac.uk/embl/ (for the HVS-I sequences [accession numbers AJ748863–AJ749247])

## References

Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phylogenet Evol 28:458–472

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB (2001) Genetic evidence on the origins of Indian caste populations. Genome Res 11:994–1004

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

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

Cavalli-Sforza LL (1997) Genes, peoples, and languages. Proc Natl Acad Sci USA 94:7719–7724

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Chen Y-S, 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 Y-S, 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, Hart WK, Renne PR, Gilbert H, Defleur A, Suwa G, Katoh S, Ludwig KR, Boisserie JR, Asfaw B, White TD (2003) Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. Nature 423:747–752

Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int J Legal Med 118:137–146

Corbo RM, Scacchi R, Rickards O, Martinez-Labarga C, De Stefano GF (1999) An investigation of human apolipoproteins B and E polymorphisms in two African populations from Ethiopia and Benin. Am J Hum Biol 11:297–304

Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331–350

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014–1022

Derbeneva OA, Starikovskaia EB, Volod'ko NV, Wallace DC, Sukernik RI (2002) [Mitochondrial DNA variation in Kets and Nganasans and the early peoples of Northern Eurasia]. Genetika 38:1554–1560

Diakonoff IM (1988) Afrasian languages. Nauka Press, Moscow

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol Biol Evol 18:672–675

Ehret C (1995) Reconstructing proto-Afroasiatic: consonants, vowels, tone and vocabulary. University of California Press, Berkeley

Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475–1484

Forster P (2004) Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. Philos Trans R Soc Lond B Biol Sci 359:255–264

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Fort M, de Stefano GF, Cambon-Thomsen A, Giraldo-Alvarez P, Dugoujon JM, Ohayon E, Scano G, Abbal M (1998) HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. Tissue Antigens 51:327–336

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345

Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JE (2002) African pastoralism: genetic imprints of origins and migrations. Science 296:336–339

Harris J (1971) The African presence in Asia: consequences of the East African slave trade. Northwestern University Press, Evanston, IL

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152–1171

Howell N, Elson JL, Turnbull DM, Herrnstadt C (2004) African haplogroup L mtDNA sequences show violations of clock-like evolution. Mol Biol Evol 21:1843–1854

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331–1334

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

Kivisild T, Rootsi S, Metspalu M, Metspalu E, Parik J, Kaldma K, Usanga E, Mastana S, Papiha S, Villems R (2003b) The genetics of language and farming spread in India. In: Bellwood P, Renfrew C (eds) Examining the farming/language dispersal hypothesis. McDonald Institute for Archaeological Research, Cambridge, UK, pp 215–222

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

Krings M, Salem A-eH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? Am J Hum Genet 64:1166–1176

Lahr M, Foley R (1994) Multiple dispersals and modern human origins. Evol Anthropol 3:48–60

Levine DN (1974) Greater Ethiopia: the evolution of multiethnic society. University of Chicago Press, Chicago

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ

(2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74:532–544

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13

Maca-Meyer N, Gonzalez AM, Pestano J, Flores C, Larruga JM, Cabrera VM (2003) Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. BMC Genet 4:15

Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249

McMahon FJ, Chen YS, Patel S, Kokoszka J, Brown MD, Torroni A, DePaulo JR, Wallace DC (2000) Mitochondrial DNA sequence diversity in bipolar affective disorder. Am J Psychiatry 157:1058–1064

Metspalu E, Kivisild T, Kaldma K, Parik J, Reidla M, Tambets K, Villems R (1999) The Trans-Caucasus and the expansion of the Caucasoid-specific human mitochondrial DNA. In: Papiha S, Deka R, Chakraborty R (eds) Genomic diversity: application in human population genetics. Kluwer Academic/Plenum Publishers, New York, pp 121–134

Militarev A (2003) Prehistory of dispersal: proto-Afrasian (Afroasiatic) farming lexicon. In: Bellwood P, Renfrew C (eds) Examining the farming/language dispersal hypothesis. McDonald Institute for Archaeological Research, Cambridge, UK, pp 135–151

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100:171–176

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992

Munro-Hay S (1991) Aksum: an African civilisation of late antiquity. Edinburgh University Press, Edinburgh

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet 62:420–434

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

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D (2003) Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. Ann Hum Genet 67:312–328

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Mehdi SQ, Torroni A, McElreavey K (2004) Where West

meets East: The complex mtDNA landscape of the southwest and central Asian corridor. Am J Hum Genet 74:827–845

Quintana-Murci L, Semino O, Bandelt H-J, 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

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet 62:531–550

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk H-V, Parik J, et al (2003) Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet 73:1178–1190

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, et al (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A (2003) Extensive female-mediated gene flow from sub-Saharan Africa into Near Eastern Arab populations. Am J Hum Genet 72:1058–1064

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

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718–726

Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-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 M-V, 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

Sauer C (1962) Seashore: primitive home of man? Proc Am Phil Soc 106:41–47

Schneider S, Roessli D, Excoffier L (2000) Arlequin version 2.000: a software for population genetics data analysis. University of Geneva, Genetics and Biochemistry Laboratory, Geneva

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034

Semino O, Santachiara-Benerecetti A, 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, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman M, Oefner P (2004) Reconstruction of patri- and matri-lineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum Mut 24: 248–260

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608

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

Stringer C (2000) Coasting out of Africa. Nature 405:24–25, 27

Tambets K, Tolk HV, Kivisild T, Metspalu E, Parik J, Reidla M, Voevoda M, Damba L, Bermisheva M, Khusnutdinova E, Golubenko M, Stepanov V, Puzyrev V, Usanga E, Rudan P, Beckmann L, Villems R (2003) Complex signals for population expansions in Europe and beyond. In: Bellwood P, Renfrew C (eds) Examining the farming/language dispersal hypothesis. McDonald Institute for Archaeological Research Monograph Series, Cambridge University Press, Cambridge UK, pp 449–458

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000) Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba—the "Black Jews of southern Africa." Am J Hum Genet 66:674–686

Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. Am J Hum Genet 70:1411–1420

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69:1348–1356

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Walter RC, Buffler RT, Bruggemann JH, Guillaume MM, Berhe SM, Negassi B, Libsekal Y, Cheng H, Edwards RL, von Cosel R, Neraudeau D, Gagnon M (2000) Early human occupation of the Red Sea coast of Eritrea during the last interglacial. Nature 405:65–69

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

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

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507