Regional Differences in the Distribution of the Sub-Saharan, West Eurasian, and South Asian mtDNA Lineages in Yemen

Viktor Černý,1* Connie J. Mulligan,2 Jakub Rídl,3 Martina Žaloudková,3 Christopher M. Edens,4 Martin Hájek,1 and Luísa Pereira5

1Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague, v.v.i., Prague 118 01, The Czech Republic
2Department of Anthropology, University of Florida, Gainesville, FL 32610-3610
3Department of Anthropology and Human Genetics, Faculty of Science, Charles University in Prague, Prague 128 00, The Czech Republic
4American Institute for Yemeni Studies, Sana’a, Republic of Yemen
5Instituto de Patologia e Imunologia Molecular da Universidade do Porto, (IPATIMUP), 4200-465 Porto, Portugal

ABSTRACT Despite its key location for population movements out of and back into Africa, Yemen has not yet been sampled on a regional level for an investigation of sub-Saharan, West Eurasian, and South Asian genetic contributions. In this study, we present mitochondrial DNA (mtDNA) data for regionally distinct Yemeni populations that reveal different distributions of mtDNA lineages. An extensive database of mtDNA sequences from North and East African, Middle Eastern and Indian populations was analyzed to provide a context for the regional Yemeni mtDNA datasets. The groups of western Yemen appear to be most closely related to Middle Eastern and North African populations, while the eastern Yemeni population from Hadramawt is most closely related to East Africa. Furthermore, haplotype matches with Africa are almost exclusively confined to West Eurasian and North African populations. In fact, Yemeni populations have the highest frequency of R0a haplotypes detected to date, thus Yemen or southern Arabia may be the site of the initial expansion of this haplogroup. Whereas three variants of the sub-Saharan haplogroup M1 were detected only in southwestern Yemen close to the Bab el-Mandeb Strait, different non-African M haplotypes were detected at low frequencies (~2%) in western parts of the country and at a higher frequency (7.5%) in the Hadramawt. We conclude that the Yemeni gene pool is highly stratified both regionally and temporally and that it has received West Eurasian, Northeast African, and South Asian gene flow.

© 2008 Wiley-Liss, Inc.
some stages of prehistory although the existence of an intercontinental bridge is unlikely during the last 80,000 years when the human intercontinental migrations occurred (Siddall et al., 2003; Forster, 2004).

The oldest archaeological evidence for modern humans on the east African coast and adaptation to marine resources comes from Eritrea ~125 KYA (Walter et al., 2000). Middle Paleolithic (MP) stone industry from the Arabian Peninsula indicates certain similarities with the MP industries of North Africa and Levant but it is not clear that all three industries are equivalent (Petraglia and Alsharekh, 2003). Archaeological findings from the Arabian Peninsula are consistent with more than one migration event (both southern and northern routes) but autochthonous technological development and diversification within Arabia cannot be excluded (Petraglia and Alsharekh, 2003).

Archaeological and historical evidence point to increasingly intense connections between Yemen and East Africa over the past 10 KYA. Chemical characterization studies imply an active early- and mid-Holocene trade in obsidian from Ethiopian and Eritrean sources into western Yemen (Francaviglia, 1995); aspects of material culture display similarities on each side of the red sea during the period about 3000–1000 BC (Fattovich, 1997; Edens, 2002; Keall, 2004). The early historic south Arabian civilization exerted considerable influence over the Ethiopian highlands around 500 BC, and Ethiopia twice occupied western Yemen, in the 3rd century and 6th centuries AD. In addition, commercial interactions connected Yemen with east Africa, and also with India and the Mediterranean basin, during the early centuries AD.

Yemen–Africa connections continued during the Islamic period. While genetic evidence points to a trade in African slaves into the Near East during the past 2,500 years (Richards et al., 2003), the role of southern Arabia remains unclear. In the reverse direction, many male Yemenis emigrated to East Africa, and also to India and southeast Asia (various papers in Freitag and Clarence-Smith, 1997; cf. Gerholm, 1977). However, despite the historical record Hudson (2003) notes that there is in fact no genetic evidence for migration of southern Arabs into Ethiopia.

During the past 3,000 years Yemen has experienced additional connections with Southwest Asia. Major linguistic, and also population, transformation began in Yemen 2,000 years ago, when Arabic-speaking groups started infiltrating from the north (Robin, 1991). A different kind of movement began late in the 9th century AD, when individual men and families started moving into Yemen from as far away as Iran. In addition, various military invasions introduced communities eventually absorbed into Yemeni populations; for example, a late pre-Islamic episode of Sassanian occupation left behind Persian communities.

The main goal of our study is to provide a description of Yemeni mitochondrial genetic variation at the regional level. Previous genetic studies of Yemen were based on only single regional samples or combined country-wide samples, precluding a regional analysis and possibly implying a genetic homogeneity at odds with the complex biological and cultural history of Yemen. Furthermore, we explicitly test hypotheses concerning the origin and timing of specific haplotypes in Yemen. We also test for evidence of gene flow into Yemen versus expansion out of Yemen. Results are discussed in relation to population history in this region of the world.

MATERIALS AND METHODS

Population samples

Buccal swabs were collected from all study participants in Yemen in 2005–2006. The selection of maternally unrelated individuals was ascertained by local Arabic-speaking assistants. The sampling strategy included only those individuals whose parents were born in the neighboring territory and excluded recent immigrants to the region. A total of 185 individuals were collected from four different geographic regions south, west, north, and east from the capital of Sana’a (see Fig. 1)—43 individuals from the southern slopes of the western highlands about 40 km southeast from Ta’izz (YTA); 67 individuals living in vicinity of Hudeida in the Tihama plain (YTI); 35 individuals living west of the Haja (or Haggia) district on the northern slopes of the western highlands (YHG); and 40 individuals from the relatively still isolated desert valley called Wadi Hadramawt in the eastern part of Yemen (YHA). For comparative purposes, we compiled previously published mtDNA HVS-I data on 37 neighboring populations in North and East Africa, the Middle East and India, focusing on studies with geographically well-defined populations (Table S1).

Laboratory analyses

DNA extractions and PCR amplification followed previously published methods (Černý et al., 2004, 2006). Amplification of HVS-I targeted nucleotides (nts) 15971–16410 with forward F-15971 (5’-TTA ACT CCA CCA TTA GCA CC-3’) and reverse R-16410 (5’-GAG GAT GGT GTG GGT CAA GGG AC-3’) primers. PCR products were purified as in Černý et al. (2004) and sequenced using the forward primer. Sequences including a homopolymeric cytosine stretch between nts 16184 and 16193 were sequenced also by the reverse primer R-16410 and length heteroplasmies were not used for subsequent statistical analyses.

Three coding region restriction fragment length polymorphisms (RFLP) that are diagnostic for basic human mtDNA phylogeny were assayed in all samples; HpaI 3594 (distinguishes L3/4/7 from L0/1/2), AluI 10400 (defines haplogroup M), and MnlI 10873 (defines haplogroup N). For further classification, three additional RFLPs were typed in a subset of individuals—AluI 7025

Fig. 1. Map of the Southern Arabia with the locations of the Yemeni samples.
Statistical analysis

Arlequin software, ver 2.00 (Schneider et al., 2000) was used for calculation of mtDNA diversity measures, such as gene diversity (Nei, 1987), nucleotide diversity (Tajima, 1983; Nei, 1987) and mean number of pair wise differences (Tajima, 1983, 1993). Departure from normal distribution of pair wise differences was tested using Harpending’s r (raggedness) index (Harpending, 1994). Tests of selective neutrality were calculated according to Tajima (1989)—D statistic and Fu (1997)—FS statistic. These parameters can be used to indicate population expansion when the null hypothesis of constant population size is rejected.

Genetic distances between Yemeni populations and populations in the comparative dataset were evaluated by $F_{ST}$ tests. Calculations were based on pair wise differences (Weir and Cockerham, 1984) and a null hypothesis of no differences between the populations was tested using 1,000 permutations. The resulting matrix of distances was visualized in two-dimensional space by means of the multidimensional scaling method (MDS) in STATISTICA, ver 5.5 (StatSoft, Inc., 2000). Analysis of molecular variance (AMOVA) (Excoffier et al., 1992) was used to investigate genetic structure among all tested populations using various groupings of populations that were based on geographic proximity and genetic distances. Two-tailed Fishers exact test P-values of $2 \times 2$ contingency tables were calculated in DnaSP (Rozas and Rozas, 1999). The reduced median (Bandelt et al., 1995) and median joining (Bandelt et al., 1999) algorithms of the network program version 4.2.0.1 (see the website fluxus-engineering.com) were used to resolve the phylogenies of numerically important haplogroups. The reduced median networks were obtained by setting the reduction threshold to the default value of 2, and these simplified data were used to obtain the median-joining networks.

RESULTS

Intra-population analyses

We assayed 185 samples (109 haplotypes) from four different regions of Yemen for mtDNA variation (HVS-I sequences and six diagnostic RFLPs) (Table S2). Summary statistics of genetic diversity in the four Yemeni groups and 37 additional populations are given in Table 1. The gene diversity ($H_{g}$) of Yemeni populations ranges from 0.955 to 0.987 with the lowest diversity in the southwest population and increasing as one moves north. Gene diversity in the eastern Yemeni populations is intermediate relative to the western populations. Generally, the Yemeni populations have lower gene diversities than the other populations (Table 1). With respect to nucleotide diversity and pair wise differences ($\pi$ and $D_{ii}$), western Yemeni populations are intermediate between the populations of East Africa (high) and Iran, Pakistan, and India (low). However, the eastern Yemeni population (YHA) has the second highest value for both $\pi$ and $D_{ii}$, and, thus, does not follow the trend of decreasing diversity as you leave east Africa. The distribution of pair wise differences in the Yemeni samples does not show any significant departure from a normal distribution (Harpending’s r index). The west coastal sample (YTI) is the only Yemeni population to exhibit a significantly negative Tajima’s D, as do more than two thirds of the 37 comparative populations, suggesting a population expansion. On the other hand, Fu’s $F_{S}$ test, which is much more sensitive (see Pereira et al., 2001), attains statistical significance in all 41 analyzed populations except the Kesra and Zriba from Tunisia (Table 1).

Inter-population analyses

The results of $F_{ST}$ tests show that the Yemeni populations are statistically distinct from each other although the western samples cluster together to the exclusion of the eastern sample (Fig. 2 shows a multidimensional scaling projection of the $F_{ST}$ matrix for all 41 populations and Table S3 contains the matrix of $F_{ST}$ tests) The west coastal (YTI) and northwestern (YHG) Yemeni populations cluster with the Bedouins of Saudi Arabia and display some affinity for the Iranian and Upper Egyptian populations, respectively. In contrast, the eastern Yemeni Hadramawt population (YHA) shows an affinity for the east African populations, particularly those from Ethiopia. It is interesting to note that, with the exception of the Assam sample which seems to be an outlier, all the Indian and Pakistani populations cluster together. In general, three main clusters can be seen in the MDS projection: 1) eastern Yemen + East Africa (Ethiopia + Kenya + Sudan), 2) western Yemen + Bedouins + North Africa (Egypt + Tunisia) + Iran, and 3) India + Pakistan (with the Assam exception as noted above).

An AMOVA provides information on the genetic structure of the populations under study (Table 2). After investigating various groupings, we concluded that the largest among-group variation (Group 3) is seen when populations are grouped according to the three clusters suggested by the MDS projection of $F_{ST}$ distances, i.e. E Africa + Middle Nile + E Yemen vs. N Africa + Iran + W Yemen + Bedouins vs. India + Pakistan. In this case, the among-group variation is the highest (5.05%) and the among-population-within-group variation is the lowest (2.83%). All variance components (Groups 1–3) are significant at the $P < 0.01$ level.

Shared HVS-I haplotypes

There is very little sharing of haplotypes across the four studied populations—99 out of 109 HVS-I haplotypes are found in only one Yemeni population (see Fig. 3 and Table S2). No haplotypes are found in all four populations and only four haplotypes are shared by three populations—one L3e3 and two R0a haplotypes are shared among the western Yemeni populations (YTA, YTI, YHG) and a L0a2 haplotype is shared by the eastern, coastal, and northwestern Yemeni populations (see Fig. 3). The eastern Hadramawt population (YHA) shows the least amount of haplotype sharing with only four haplotypes (13.8%) appearing in another Yemeni population, i.e. sub-Saharan haplotypes L2a1b, L3f1, and L0a2 are shared with the coastal Tihamah (L0a2 is also found in northwestern Hajja, YHG) and a J1c haplotype also occurs in the southwestern Ta’izz sample (YTA). Of the western populations, Ta’izz shows the highest frequency...
of haplotype sharing (25.9%; one L3e3, two R0a, one J1c, and one J*)). Tihama (YTI) shows slightly less sharing (23.0%; one L0a2, one L2a1b, one L3e3, one L3f1, two R0a, one J1c, and one J*) and Haja shares only four haplotypes (14.3%; one L0a2, one L3e3, and two R0a haplotypes) with other Yemeni populations.

The Yemeni populations also share some haplotypes with East and Northeast African populations, as represented in several datasets (Watson et al., 1997; Krings et al., 1999; Kivisild et al., 2004) (Table S4a-d). Almost all haplotypes matches between southwestern Yemen (YTA) and Africa belong to West Eurasian R0a haplogroup, previously known as (pre-HV1) (Table S4a). In particular, the Yemeni haplotype defined by mutations 16126–16362 is found to be very common in all four Ethiopian populations (Tigrai, Amhara, Oromo, Gurage) and in the Egyptians from Upper Egypt. The Tigrais, Amhara, and Egyptians from Lower Egypt also share some West Eurasian J and T haplotypes with the Ta'izz population. It should be noted that no matches with African L-type haplogroups were found between the Ta'izz population and any of the African populations.

In west coastal Yemen, the matches between the Tihama population (YTI) and Africa are predominantly in the R0a haplogroup, although not as pronounced as in the southwestern Yemeni population (Table S4b). The most frequent haplotype in west coastal Yemen is 16126–16362, which is found not only in the Ethiopian highlands but also in Somalia, lower Egypt and at especially high frequency in the Nubians. The Tihama share some West Eurasian haplotypes with Africans, e.g. J and K with Ethiopians, Somali, and Egyptians. Two sub-Saharan haplotypes (L0a2 and L3f1) are also shared between the Tihama and Turkan, Kikuyu (L02a) and Tigrai, Amhara, Somali, and Nubians (L3f1).

Haja, the last western Yemeni population (YHG), shares the most diverse set of haplotypes with Africans, of which R0a haplogroup represents approximately half of the matches (Table S4c). Matches are also found with West Eurasian haplotypes H (Tigrais and Upper Egyp-
Eastern Yemen is quite different than western Yemen in terms of haplotype matches with Africans (Table S4d). Most significantly, no R0a haplotype is shared between the Hadramawt population (YHA) and Africans. The only matches with European haplotypes are found in the Hadramawt population (YHA) and Africans. The only matches with European haplotypes are found in the Amhara with single matches to J1c and HVI. On the other hand, numerous matches were identified with L-type haplogroups. L3f1 matches are particularly numerous and are found in almost all Ethiopian samples (except the Oromos), Somali and Nubians. Multiple L0a2 matches are found in the Turkana and Kikuyu and a single L0a1a match is found in the Kikuyu. For the first time in the Yemeni, a match is found for L3d1 in the Oromos. Furthermore, the eastern Yemeni are distinguished from the western Yemeni by sharing several M and N haplotypes with Africans; N1a (Tigrais, Amharas, Gurage) and M30 (Upper Egypt and Kesra).

Phylogeography

A sub-Saharan origin can be demonstrated in one third (34.1%) of the mtDNA haplogroups detected in four sampled Yemeni populations (overall proportions of major mtDNA clades in four Yemeni populations are shown in Fig. 4). Averaged frequencies (averaged over four Yemeni populations) of the sub-Saharan mtDNA haplogroups are 9.2% for L0a, 2.2% for L1b’c, 5.4% for L2, 16.2% for L34 (distinguished by +HpaI 3594) and 1.1% for M1. The geographic origin of the remaining mtDNA haplogroups (J, T, N, R and most of U, overall 62.7%) can be situated in Western Eurasia and, eventually, in South Asia (most of M and possibly also U7, 3.2%). Whereas a gradual south-to-north increase in sub-Saharan mtDNA haplogroups can be observed in the western populations (Ta’izz-16.3%, Tihama-28.4% Hajja-34.3%), the frequency of sub-Saharan haplogroups is significantly higher in Hadramawt (60.0%) (two-tailed Fisher’s exact test P-value are 0.00044 and 0.002054 with Ta’izz and Tihama, respectively, while the P-value for comparison with Hajja, 0.037195, is not significant after Bonferroni correction).

A more detailed analysis of the regional distribution of mtDNA haplogroups in Yemen is presented in Table 3. Most sub-Saharan L haplotypes are present in only 1–3 individuals per population (1.5–8.6%) with the exception of seven L0a2 individuals in Hadramawt (17.5%). L0a haplotypes (L0a*, L0a1a, L0a1, and L0a2) are the most frequent sub-Saharan mtDNA haplogroup and are found in almost all populations (except southwestern Yemen) at frequencies of 1.5–17.5%.

Figure 5 shows a minimum spanning network of the L0a haplotypes with the Yemeni haplotypes noted. The two haplotype closest to the L0a root haplotype are HPT3 and HPT1, which are one and two steps away and are found in Hajja and Tihama, respectively (Fig. 5a). The others haplotypes belong to L0a1 and are more derived sequences found in Hajja and/or in Hadramawt. On the other hand, the most widespread L0a2 haplotype in Yemen (HPT8, 9 individuals) is the root haplotype (Fig. 5b) with many exact matches in Southeast Africa (Pereira et al., 2001; Salas et al., 2002).

L1b and L1c haplotypes are quite rare in Yemen, occurring in only single individuals in Hajja (L1b), and Ta’izz (L1c2), and two individuals in Hadramawt (L1c1). L2a haplotypes (L2a, L2a1, L2a1a, and L2a1b) are found in only single individuals throughout all four Yemeni populations. Haplogroup L2b is found in only two individuals in Ta’izz. Haplogroup L3 is represented by many haplotypes, including L3a, L3b, L3d1, L3d2, L3e1a, L3e2b, L3c3, L3f*, L3f1, L3h, and L3w, that are distributed throughout all four Yemeni populations with the highest frequencies found in Hadramawt (27.5% total). Only one L4 haplotype (L4g) was detected in Tihama.

In contrast to the high frequency of sub-Saharan haplotypes in the eastern Hadramawt, western Yemeni populations exhibit West Eurasian mtDNA haplotypes as their most frequent haplotype. For instance, haplotype R0a is the most common haplotype in Ta’izz (25.6%) and

![Fig. 2.](image-url) The multidimensional scaling projection of the FST matrix, for abbreviations see Table S1.

### TABLE 2. Proportions of variation in different population groupings (AMOVA)

<table>
<thead>
<tr>
<th></th>
<th>Among groups</th>
<th>Among populations (within groups when grouping)</th>
<th>Within populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>All populations</td>
<td>5.88</td>
<td>94.12</td>
<td>94.12</td>
</tr>
<tr>
<td>Group 1 (E. Africa + Arabian peninsula + India vs. NE Africa + Iran)</td>
<td>2.16</td>
<td>5.03</td>
<td>92.75</td>
</tr>
<tr>
<td>Group 2 (E. Africa + E. Yemen + India vs. NE Africa + Iran + W. Yemen + Bedouins)</td>
<td>2.89</td>
<td>4.60</td>
<td>92.51</td>
</tr>
<tr>
<td>Group 3 (E. Africa + Middle Nile + E. Yemen vs. NE. Hadramawt + W. Yemen + Bedouins vs. India + Pakistan)</td>
<td>5.05</td>
<td>2.83</td>
<td>92.12</td>
</tr>
</tbody>
</table>
Hajja (22.9%). In west coastal Yemen (Tihama), J1 and J1c are the most frequent haplotypes at frequencies of 10.4 and 11.9, respectively.

Haplogroup M, which has been proposed as a marker of the initial migration of anatomically modern humans out of Africa, is also distributed unequally throughout the Yemeni study populations. Two M1 and a single M4 haplotype are found in southwestern Yemen (7.0%), a single M40 haplotype in west coastal Yemen, and M2b, M3, and M30 haplotypes in Hadramawt (7.5%). No M haplotypes are found in northwest Yemen.

**DISCUSSION**

**Regional differentiation**

We present the first regional analysis of Yemeni mitochondrial genetic diversity. Three populations in western Yemen (two highland and one coastal population) and one population in the eastern desert were assayed for mitochondrial control region DNA sequence and coding region variants. This region of southern Arabia has been the site of multiple migrations originating from Africa to southwest Asia and back over the past tens of thousands of years. Our results demonstrate geographic differentiation within the Yemeni populations that likely reflects gene flow from different source populations in Africa, West Eurasia, and South Asia. MDS plots of genetic distances demonstrate that western populations are most closely related to Middle Eastern and North African populations while the eastern population is most closely related to East Africans. Furthermore, haplotype matches with Africans are almost exclusively confined to the West Eurasian R0a, previously known as (pre-HV)1, haplogroup in southwestern Yemen (YTA) although more matches with sub-Saharan L-type haplogroups appear in more northern populations (YTI, YHG). In contrast, no R0a haplotype matches are found in eastern Yemen (YHA), where virtually all matches are with sub-Saharan L or South Asian M or N haplogroups.

**Migration footprints**

Our results shed light on the first migration of anatomically modern humans out of Africa and subsequent migrations between Africa and Arabia. Our intense sampling of Yemeni populations reveals the highest frequency of sub-Saharan L haplotypes detected in Arabia to date (ranging from 16.3% in southwestern Yemen to 60% in eastern Hadramawt), attesting to the close ties between Africa and southern Arabia. The eastern Hadramawt, located ~800 km east of Africa, display increased levels of nucleotide diversity relative to western Yemeni populations, thus defying an expectation of decreased diversity with increasing distance from Africa. Furthermore, the Hadramawt carry the highest levels of sub-Saharan haplotypes and cluster most closely with East African populations. In contrast, despite the geographic proximity of western Yemen and East Africa and cultural contacts between these regions dating from at least the 7th millennium BC, our samples from western Yemen cluster most closely with Middle Eastern and North African populations. Indeed, western Yemeni populations carry the highest frequency reported to date of the ancestral types of the European R0a haplogroup (as compared with Abu-Amero et al. (2007), Fig. 3). Richards et al. (2000, 2003) have proposed a Middle Eastern origin for haplogroup (pre-HV)1/R0a, thus Yemen may be the site of the initial expansion of this haplogroup, which eventually evolved into haplogroup H, the predominant haplotype in Europe. It is interesting to note that previous genetic studies of Yemeni populations failed to detect a high frequency of R0a, perhaps because of limited sampling (Richards et al., 2000; Thomas et al., 2002; Kivisild et al., 2004; although Richards et al., 2003 did find four R0a/(pre-HV)-1 individuals in their sample of 56 Yemen Hadramawts).

A relevant question is whether or not the observed affinity of Hadramawt and East Africa reflects an ancient migration of humans along the “southern dispersal route” out of Africa. It has been suggested that haplogroup M may carry a signature of this ancient migration event (Quintana-Murci et al., 1999; Forster and Matsumura, 2005; Macaulay et al., 2005). Our results show that M haplogroup is represented at low frequencies in Yemen, ranging from zero in northwestern Yemen to 7.5% in Hadramawt. However, the M haplotypes in Hadramawt have exact or one step matches with Indian M haplotypes (Kivisild, personal communication) suggesting a non-African and non-Arabian origin. Richards et al. (2003) has interpreted the high frequency of L lineages in Yemeni Hadramawt as evidence of gene flow due to the Arab slave trade ~2,500 years ago. Our southwestern Yemeni sample, closest to the Bab el-Mandeb Strait, is the only population that carries the sub-

---

American Journal of Physical Anthropology
Saharan M1 haplotype \((n = 2)\), which arose in southwestern Asia from where it back-migrated to Africa \(40–45\) KYA (Olivieri et al., 2006). The geographic origin of diversification of the M clades is relevant to the evolutionary history of the entire region of East/Northeast Africa, Arabia, and southwest Asia, yet this issue is unresolved. Higher diversity values and older coalescent times are evidenced in India M haplotypes (Quintana-Murci et al., 1999; Metspalu et al., 2004; Rajkumar et al., 2005; Thangaraj et al., 2006), but the coalescence time for the Ethiopian M haplogroup is only about 12,000 years younger with more or less overlapping confidence intervals (Quintana-Murci et al., 1999; Kivisild et al., 2004). Thus, it is possible that more intense sampling of southern Arabia (Yemen, Oman) and neighboring regions in Africa (Eritrea, Djibouti) could reveal novel ancestral M lineages that could clarify the itinerary of the first successful out-of-Africa migration.

**Macrohaplogroup L**

The approximate time period(s) of sub-Saharan influences in Yemen can be investigated via the distribution of L-haplotypes in a phylogeographic perspective. L0a

<table>
<thead>
<tr>
<th>Haplogroup</th>
<th>n ALL</th>
<th>% ALL</th>
<th>n YTA</th>
<th>% YTA</th>
<th>n YTI</th>
<th>% YTI</th>
<th>n YHG</th>
<th>% YHG</th>
<th>n YHA</th>
<th>% YHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0a*</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0a1</td>
<td>2</td>
<td>1.1</td>
<td>2</td>
<td>5.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0a1a</td>
<td>3</td>
<td>1.6</td>
<td>1</td>
<td>2.9</td>
<td>2</td>
<td>5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0a2</td>
<td>11</td>
<td>5.9</td>
<td>3</td>
<td>4.5</td>
<td>1</td>
<td>2.9</td>
<td>7</td>
<td>17.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1b</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>5.0</td>
</tr>
<tr>
<td>L1c1</td>
<td>2</td>
<td>1.1</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1c2</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>5.0</td>
</tr>
<tr>
<td>L2a</td>
<td>2</td>
<td>1.1</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
<td>2.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2a1</td>
<td>3</td>
<td>1.6</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2a1a</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2a1b</td>
<td>2</td>
<td>1.1</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2b</td>
<td>2</td>
<td>1.1</td>
<td>2</td>
<td>4.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3a</td>
<td>2</td>
<td>1.1</td>
<td>2</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3b</td>
<td>4</td>
<td>2.2</td>
<td>3</td>
<td>4.5</td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3d1</td>
<td>3</td>
<td>1.6</td>
<td>1</td>
<td>2.3</td>
<td></td>
<td></td>
<td>3</td>
<td>7.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3d2</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>5.0</td>
</tr>
<tr>
<td>L3e1a</td>
<td>3</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>7.5</td>
</tr>
<tr>
<td>L3e2b</td>
<td>2</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>L3e3</td>
<td>7</td>
<td>3.8</td>
<td>2</td>
<td>4.7</td>
<td>2</td>
<td>3.0</td>
<td>3</td>
<td>8.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3f*</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3f1</td>
<td>4</td>
<td>2.2</td>
<td>2</td>
<td>3.0</td>
<td>2</td>
<td>5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3h</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>L3w</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>J4</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td></td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1a</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td>1</td>
<td>2.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1c</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1d</td>
<td>1</td>
<td></td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>2</td>
<td>1.1</td>
<td></td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R0</td>
<td>24</td>
<td>13.0</td>
<td>11</td>
<td>25.6</td>
<td>4</td>
<td>6.0</td>
<td>8</td>
<td>22.9</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>R2</td>
<td>2</td>
<td>1.1</td>
<td></td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>9</td>
<td>4.9</td>
<td>6</td>
<td>9.0</td>
<td>1</td>
<td>2.9</td>
<td>2</td>
<td>5.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HV1</td>
<td>5</td>
<td>2.7</td>
<td>5</td>
<td>3.0</td>
<td>1</td>
<td>2.9</td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J1</td>
<td>11</td>
<td>5.9</td>
<td>3</td>
<td>7.0</td>
<td>7</td>
<td>10.4</td>
<td>1</td>
<td>2.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J1b</td>
<td>6</td>
<td>3.2</td>
<td>1</td>
<td>2.3</td>
<td>2</td>
<td>3.0</td>
<td>1</td>
<td>2.9</td>
<td>2</td>
<td>5.0</td>
</tr>
<tr>
<td>J1c</td>
<td>12</td>
<td>6.5</td>
<td>3</td>
<td>7.0</td>
<td>8</td>
<td>11.9</td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2a</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td>1</td>
<td>2.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5</td>
<td>2.7</td>
<td>3</td>
<td>7.0</td>
<td>1</td>
<td>2.9</td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T5</td>
<td>3</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.9</td>
<td>1</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>7</td>
<td>3.8</td>
<td>4</td>
<td>9.3</td>
<td>3</td>
<td>4.5</td>
<td></td>
<td>1</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>U1a</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U4</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U5a1a</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td>1</td>
<td>2.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U7</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U8b</td>
<td>2</td>
<td>1.1</td>
<td>2</td>
<td>4.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>2</td>
<td>1.1</td>
<td>2</td>
<td>4.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2b</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>M40</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>M30</td>
<td>1</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|         | 185   | 100.0 | 43    | 100.0 | 67    | 100.0 | 35    | 100.0 | 40    | 100.0 |
haplotypes are the most common L sequences in Yemen, ranging from 0 in Ta‘izz to 22.5% in Hadramawt. The Yemeni sequences are present as ancestral types (L0a1 in Hajja and L0a2 in Hadramawt) and as various derivatives of the ancestral sequences, thus they also represent a wide range of diversity within the Yemeni populations. Presence of these haplotypes in Yemen might represent recent gene flow from the Arabian Peninsula, particularly from the Yemeni Highlands. The presence of these haplotypes in Yemen is consistent with the East African distribution of mtDNA clades, particularly M haplogroups.

CONCLUSIONS

Our study provides evidence of multiple contacts between eastern Yemen and Africa. However, southwest Yemen, which is closest to East Africa, has the lowest frequency of African L haplotypes. Instead, Middle Eastern and Eurasian genetic influences appear to predominate in western Yemen. It is difficult to determine the antiquity and directionality of these genetic exchange events suggested throughout Yemen. There is support for multiple sources of incoming gene flow (various L lineages from East, West, North, and Southeast Africa and Middle Eastern lineages from India), and also indication of an expansion event out of Yemen (R0a/prior-HV1). Additional fine-scale analysis of Yemeni populations is necessary to fully explore the role of southern Arabia in the myriad migrations that have occurred throughout human history at this crossroads of three continents. Specifically, Y chromosome and autosomal variation should be investigated to expand the mitochondrial results and to test hypotheses of sex-biased demographic processes as have already been found in Africa and Arabia (Richards et al. 2003; Deströ-Bisol et al. 2004; Wood et al. 2005).

LITERATURE CITED


American Journal of Physical Anthropology