Out of Arabia—The Settlement of Island Soqotra as Revealed by Mitochondrial and Y Chromosome Genetic Diversity

Viktor Černý,[1]* Luísa Pereira,[2] Martina Kujanová,[3] Alžběta Vašíková,[4] Martin Hájek,[5] Miranda Morris,[6] and Connie J. Mulligan[7]

[1]*Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague 118 01, The Czech Republic*
[2]*Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), 4200-465 Porto, Portugal and Medical Faculty, University of Porto, 4200-319 Porto, Portugal*
[3]*Department of Anthropology and Human Genetics, Faculty of Science, Charles University in Prague, Prague 128 00, The Czech Republic*
[4]*Institute of Hematology and Blood Transfusion, Prague 128 20, The Czech Republic*
[5]*Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague 118 01, The Czech Republic*
[6]*Honorary Research Fellow, School of History, University of St. Andrews, St. Andrews, Scotland*
[7]*Department of Anthropology, University of Florida, Gainesville, FL 32610-3610*

KEY WORDS mtDNA and NRY diversity; regional sampling; phylogeography; migrations; Southern Arabia

ABSTRACT The Soqotra archipelago is one of the most isolated landmasses in the world, situated at the mouth of the Gulf of Aden between the Horn of Africa and southern Arabia. The main island of Soqotra lies not far from the proposed southern migration route of anatomically modern humans out of Africa ~60,000 years ago (kya), suggesting the island may harbor traces of that first dispersal. Nothing is known about the timing and origin of the first Soqotri settlers. The oldest historical visitors to the island in the 15th century reported only the presence of an ancient population. We collected samples throughout the island and analyzed mitochondrial DNA and Y-chromosomal variation. We found little African influence among the indigenous people of the island. Although the island population likely experienced founder effects, links to the Arabian Peninsula or southwestern Asia can still be found. In comparison with datasets from neighboring regions, the Soqotri population shows evidence of long-term isolation and autochthonous evolution of several mitochondrial haplogroups. Specifically, we identified two high-frequency founder lineages that have not been detected in any other populations and classified them as a new R0a1a1 subclade. Recent expansion of the novel lineages is consistent with a Holocene settlement of the island ~6 kya. Am J Phys Anthropol 000:000–000, 2009.

The four islands of the Soqotra archipelago lie in the Gulf of Aden, a short detour from the main sea routes that crisscross the Indian Ocean linking the Red Sea, India, Arabia, and East Africa. The archipelago was once part of the Gondwanaland supercontinent but became isolated from Africa, Arabia and India when the landmasses separated in the Cretaceous, at least 60 million years ago. The archipelago is part of the major tectonic plate of Continental Africa, with the deep Gulf of Aden carrying the main, ever-widening fault that separates the Arabian and African plates.

The largest island, Soqotra, lies some 380 km south of Ras Fartak on the Arabian mainland, and covers an area of some 3650 sq km., with the granite peaks of its central mountains rising to a height of 1519 m. The population of some 50,000 people subsists mainly on fishing, the cultivation of date-palms and pastoralism. Long-term isolation explains the diversity and uniqueness of the Soqotran flora and fauna and also accounts for the myths and legends which have enveloped the island from earliest times. Of the 825 plant species, over a third are endemic (Cheung and DeVantier 2006), many of them remnants of ancient flora that disappeared long ago from the African-Asian mainland. The central mountains alone contain over 200 endemic species. The climate of the archipelago is monsoonal, and heavily influenced by the major wind and water circulation movements within the Arabian Sea and Indian Ocean. The

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Council of American Overseas Research Centers; American Institute for Yemeni Studies; Programa Operacional Ciência, Tecnologia e Inovação (POCTI); Quadro Comunitário de Apoio III; Ministry of Education of the Czech Republic, Grant number: KONTAKT ME 917; United States National Science Foundation, Grant number: BCS-0518530.

*Correspondence to: Viktor Černý, Archaeogenetics Laboratory, Institute of Archaeology, Letenska 4, 118 01, Prague 1, The Czech Republic. E-mail: cerny@arup.cas.cz

Received 24 June 2008; accepted 24 September 2008

DOI 10.1002/ajpa.20960
Published online in Wiley InterScience (www.interscience.wiley.com).

hot, dry winds of the south-west monsoon, which lasts from May to September, effectively close the island to sea traffic and fishing. The fact that the islanders owned no sea-going boats until relatively recently (early 20th century) and had to rely on the seasonal visits of trade-boats for many of their necessities only intensified their isolation.

Unfortunately, prehistory and origins of the aboriginal Soqotri population are still not well understood as only relatively late historical events are known. In 1480, Soqotra was ruled by sultans of the Al 'Afrariya Sultanate of the Mahra on the Yemeni mainland (Serjeant, 1963). With a brief interlude (1507–1511) when the Portuguese set up their own garrison, the Mahra Sultans ruled until 1967, when control of the island passed to the new government of the Peoples Democratic Republic of Yemen (PDRY) that took control of southern Arabia when the British left Aden. After the unification of North and South Yemen in 1990, Soqotra became part of the Republic of Yemen. In the 1890s, when the seat of the Mahra Sultanate transferred to Soqotra, many Mahra tribesmen came to settle on the island. Others who chose to make their home on Soqotra were Arab merchants and middlemen from the Gulf of Aden, the Hadramawt governate in Yemen and Oman, and members of the Saadah family from the Hadramawt (a group accredited with special spiritual powers and descent from the family of the Prophet Muhammad). Less willing settlers were slaves from the African mainland who labored for the above.

It is believed on the island that over time the aboriginal Soqotrans were driven away from the lusher areas of the island (areas of superior pasture where water and soil permitted the cultivation of finger-millet and date-palms). Thus the well-watered plains and the northern foothills of the Haghier, the valleys lying south of the same mountains, and the more accessible parts of the mountains became owned or controlled by the incoming groups. Oral history and tradition associates this with the arrival of greater numbers of powerful Mahra (and associated tribes) once their Sultan began to rule and tax-farm the island (15th century onwards) but also with depredations of pirates in the region over the centuries (exact period unknown). The original mountain dwellers became confined to the drier fringes of the mountains and plateaus and the more inaccessible areas.

The inhabitants of Soqotra speak Soqotri, a unique South Arabian language within the Semitic language family. The island of Soqotra can be roughly divided into geographic areas in which different dialects of Soqotri are spoken: the eastern end of the island, in and around the Haghier mountains, the western end of the island, and the settled towns and villages of the northern coast. Such divisions are also relevant ethnographically (Morris, 2002).

Although several studies have been conducted on genetic variation of Soqotri flora and fauna, human genetic variation on the island has never been investigated. In fact, there is no clear hypothesis for the initial settlement of the island, which lies not far from the proposed southern migration route of anatomically modern humans out of Africa (Metspalu et al., 2004; Macaulay et al., 2005) suggesting the island may harbor traces of the first Exodus. We address two main research questions: i) when was the island Soqotra first settled? and ii) from where did the first settlers originate? These questions were investigated by a study of two types of nonrecombining DNA loci that have been demonstrated to be very advantageous for such kind of study (Rando et al., 1999; Thangaraj et al., 2005). We collected a sample of 65 indigenous Soqotri individuals living throughout the island and we provide the genetic analysis of mitochondrial DNA (mtDNA) and the nonrecombining portion of the Y chromosome (NRY).

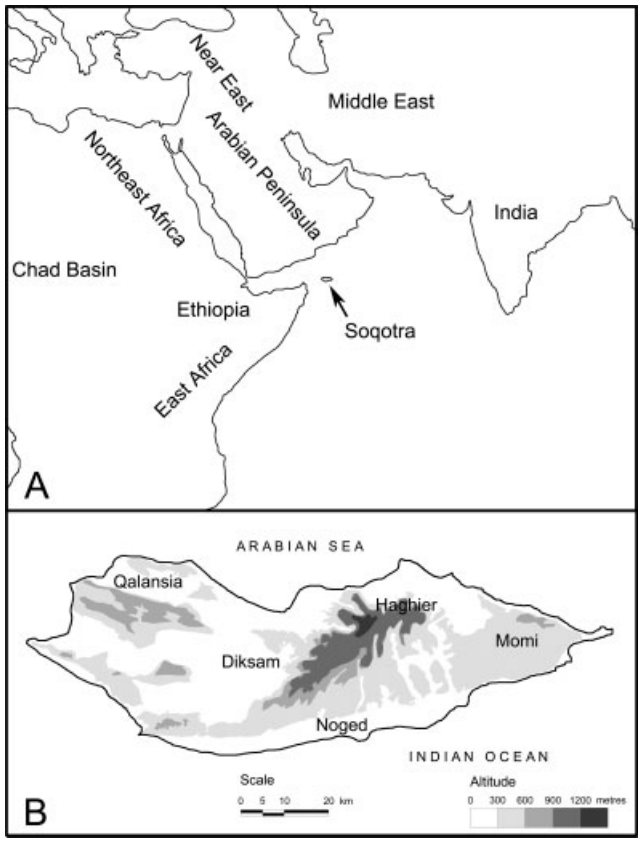

Fig. 1. (**A**) Map of the Horn of Africa and Arabian Peninsula including the island of Soqotra. (**B**) Enlargement of Soqotra with sample collection sites marked.

MATERIALS AND METHODS

Population samples

The Soqotri population was sampled throughout the island while avoiding the capital city of Hadiboh where many recent immigrants live. We tried to secure the samples from all main areas of the island. The samples collected covered the population of the high limestone plateaus beside the Haghier mountains (Diksam, $n = 9$), the Haghier mountains ($n = 7$), the western capital of Qalansiyah ($n = 13$), and Momi ($n = 6$) in the east. We also collected samples from the population of the southern plain (Noged, $n = 22$), the eastern end of which is inhabited by people largely of eastern origin while the western end was more recently settled (since the 1970s) by people whose place of origin is the western highlands. The remaining eight individuals were from very small hamlets distributed through the island. We gathered information on both maternal and paternal ancestry of examined individuals to avoid inclusion of related indi-

TABLE 1. Diversity indices for mtDNA and NRY in Soqotra and neighboring populations

pop (mtDNA)	*n*	*k*	$H_s \pm$ SE	$\pi \pm$ SE	D_{ii}	$V(D_{ii})$	*r*	*p(r)*
YSO	65	17	0.896 ± 0.020	0.015 ± 0.008	5.07	9.10	0.021	0.48
YTI	67	39	0.974 ± 0.008	0.020 ± 0.011	6.89	10.60	0.012	0.26
SY1	69	58	0.991 ± 0.006	0.017 ± 0.009	5.69	7.04	0.009	0.77
PER	42	38	0.992 ± 0.009	0.017 ± 0.009	5.79	6.19	0.009	0.87
GUJ	53	49	0.997 ± 0.004	0.018 ± 0.010	6.26	5.52	0.011	0.54
EG2	58	49	0.993 ± 0.005	0.024 ± 0.013	8.31	12.78	0.004	0.88
TIG	53	46	0.994 ± 0.005	0.024 ± 0.013	8.24	9.30	0.009	0.34
TUK	37	34	0.994 ± 0.009	0.030 ± 0.016	10.20	13.23	0.007	0.69
KOT	56	31	0.955 ± 0.017	0.021 ± 0.011	7.21	17.41	0.016	0.21
pop (NRY)	*n*	*k*	$H_s \pm$ SE					
YSO	63	5	0.260 ± 0.070					
QAT	72	10	0.635 ± 0.059					
YEM	62	5	0.450 ± 0.070					
OMA	120	12	0.783 ± 0.026					
MEA	23	6	0.791 ± 0.054					
INO	80	9	0.636 ± 0.050					
SUD	40	3	0.554 ± 0.059					
ETH	88	4	0.450 ± 0.050					
SOM	201	7	0.299 ± 0.040					

pop, population sample (for codes see Supp. Info. Table 1 for mtDNA and Table 2 for NRY); *n*, sample size; *k*, number of haplotypes; $H_s \pm$ SE, gene diversity ± standard error; π, nucleotide diversity ± standard error; D_{ii}, mean number of pairwise differences (mismatch observed mean); $V(D_{ii})$, mismatch observed variance; *r*, Harpending's raggedness index; *p*(*r*), probability of Harpending's index; for more population samples tested, see Černý et al., 2008.

viduals. The collection sites are shown on the map of island Soqotra (see Fig. 1). A total of 65 individuals were screened for mtDNA hypervariable segment I (HVS-I) diversity and mitochondrial coding region single nucleotide polymorphisms (SNPs). A subset of 63 individuals was assayed for Y-chromosome SNPs.

Laboratory analyses

DNA extractions and PCR amplifications of mtDNA HVS-I were carried out according to published protocols (Černý et al., 2004; 2006). Amplicons were purified by QIAquick PCR Purification Kit and sequenced using the forward amplification primers. In some cases, e.g. the poly-C stretch between nt 16184–16193, the reverse primer was used for sequencing. The sequencing reactions were electrophoresed on a 3100 DNA Sequencer (Applied Biosystems, Forster City, CA). Chromatograms were evaluated using BioEdit software version 7.0.4.1 (Hall, 1999). In cases of ambiguous sequencing results, new amplifications and sequencing reactions were performed. Diagnostic coding region SNPs were used to identify the main branches of the human mtDNA phylogeny. Specifically, 3594*Hpa*I, 10400*Alu*I, 10873*Mnl*I, and 4216*Nla*III were used to distinguish haplogroup L3 from all L sub-Saharan haplogroups, haplogroup M from L3, haplogroup N from L3, and haplogroup JT from R, respectively. The haplogroup N was distinguished from R by direct sequencing of the region around nt12705 (Krings et al., 1999; Salas et al., 2002; Kivisild et al., 2004; Torroni et al., 2006).

For classification of Y chromosome haplogroups, the Signet Y-SNP Identification System v 2.0 (Marligen) was used. We analyzed 10 binary SNP markers by A-R multiplex (M9, M45, M89, M96, M122, M168, M175, M207, M304, and M343) that divided the samples according to the general phylogenetic classification of the Y chromosome (Karafet et al., 2008). Subsequently, as the majority of the samples belonged to haplogroup J, we assayed additional SNPs, such as M47, M67, M92, M172, M241, and M267, which have been used for more detailed classification of this haplogroup.

Statistical and phylogenetic analyses

Mitochondrial DNA and NRY data were used to investigate genetic diversity in the Soqotri population using Arlequin 3.0 (Excoffier and Schneider 2005). Particularly, we analyzed gene diversity for both mtDNA and NRY data and nucleotide diversity, mismatch distributions (Tajima, 1983; Nei, 1987; Tajima 1993) and Harpending's raggedness index (Harpending, 1994) for mtDNA data.

For population comparisons, raw sequence data were used for mitochondrial diversity. However, for the Y chromosome, it was necessary to pool some haplogroups in some populations for a final comparison of only 12 haplogroups that were present in all datasets.

We calculated F_{ST} distances using Arlequin 3.0. We used mtDNA and NRY data from various populations for comparative purposes (see Supp. Info. Tables 1 and 2 for mtDNA and NRY data, respectively). Subsequently, genetic distances between the populations were visualized using Multidimensional Scaling (MDS) by means of the SPSS 10.0 software (SPSS Inc, Chicago, IL).

MtDNA haplotypes were classified into major haplogroups according to the classification system presented in previous publications (Salas et al., 2002; Reidla et al., 2003; Kivisild et al., 2004; Palanichamy et al., 2004; Achilli et al., 2005; Bandelt et al., 2006; Olivieri et al., 2006; Torroni et al., 2006; Roostalu et al., 2007; Behar et al., 2008). Phylogenetic analysis of mtDNA haplotypes was performed using the reduced median algorithm (Network 4.5.0.0), with a reducing factor of two (Bandelt et al., 1995), followed by the median joining algorithm to resolve intermediate nodes. For calculation of time to most recent common ancestor (TMRCA), ρ statistics (mean divergence from inferred ancestral haplotype) were used with a HVS-I mutation rate of one transition per 20,180 years (Forster et al., 1996). The standard

TABLE 2. Shared mtDNA haplotypes between Soqotra and neighboring regions

hpt	HVS-I	a	b	c	d	e	f	HG	Soqotra ($n = 65$)	Arabia ($n = 619$)	N East ($n = 767$)	M East ($n = 1198$)	N India ($n = 854$)	S India ($n = 1436$)	NE Africa ($n = 344$)	Ethiopia ($n = 301$)	E Africa ($n = 674$)	Chad B ($n = 448$)
1	111 177 184 223 301 304	–	–	–	–	#	#	L3*	1; 1.2[a]									
2	111 184 223 304	–	–	–	–	#	#	L3*	2; 3.1							2; 0.7		
3	086 129 148 185 223	–	–	+	–	#	#	N*	1; 1.2									
4	086 129 148 223	–	–	+	–	#	+	N*	15; 23.1									
5	147G 172 223 248 355	–	–	+	–	+	#	N1a	4; 6.2	2; 0.3						6; 2.0	1; 0.1	
6	362	–	–	+	–	#	+	R*	1; 1.2	2; 0.3	6; 0.8	9; 0.8			5; 1.5			
7	rCRS	–	–	+	–	#	–	H	2; 3.1	8; 1.3	48; 6.3	54; 4.5	14; 1.6	11; 0.8	10; 2.9	1; 0.3		5; 1.1
8	126 172 180 355 362	–	–	+	–	#	#	R0a1b	1; 1.2									
9	126 172 292 355 362	–	–	+	–	#	#	R0a1b	1; 1.2									
10	126 172 355 362	–	–	+	–	–	#	R0a1b	10; 15.4									
11	126 355 362	–	–	+	–	#	+	R0a1	4; 6.2	19; 3.1	4; 0.5	3; 0.3			1; 0.3	7; 2.3		
12	126 362	–	–	+	–	#	+	R0a	9; 13.8	10; 1.6	3; 0.4	1; 0.1	4; 0.5		6; 1.7	2; 0.7	1; 0.1	
13	069 126	–	–	+	+	#	+	J*	3; 4.6	5; 0.8	6; 0.8	4; 0.3			2; 0.6	1; 0.3		
14	069 126 179	–	–	+	+	#	+	J*	3; 4.6	1; 0.2								
15	069 126 145 222 256 261 278	–	–	+	+	#	#	J1b	2; 3.1									
16	126 146 189 294 296	–	–	+	+	#	#	T*	1; 1.2									
17	126 294 296 362	–	–	+	+	–	+	T2	5; 7.7			1; 0.1						
TSS										47; 7.6	67; 8.7	72; 6.0	18; 2.1	11; 0.8	24; 7.0	19; 6.3	2; 0.3	5; 1.1

[a] Absolute and relative frequencies in percentage after semicolon.
Coding region SNPs are reported here (a, 3592 *Hpa*I; b, 10397 *Alu*I, 10871 *Mnl*I; d, 4216 *Nla*III; e, 12705; f, 7025 *Alu*I) but the analysis of shared haplotypes was based on HVS-I sequences only. Unique Soqotri haplotypes are shaded in gray.
TSS, total sequences shared.

deviation of the ρ estimator was calculated according to Saillard et al. (2000).

RESULTS

Intrapopulation diversities

Diversity indices for mtDNA and NRY data are presented in Table 1. Based on mtDNA HVS-I sequence and coding region SNPs, we identified only 17 different mtDNA haplotypes among 65 Soqotri samples (Table 1 for mtDNA). The gene and nucleotide diversities (H_s) as well as mean number of pairwise differences (D_{ii}) were rather low when compared with a set of various neighboring populations of similar sample sizes included in Supporting Information Table 1. Harpending's raggedness index did not show significant values in any sample analyzed. Similar results have been achieved for NRY data. For these comparisons we used the same level of resolution of the phylogeny of NRY haplogroups in all the compared populations what can influence the diversity levels of more distant groups. Nevertheless, the Table 1 (for NRY) shows that similarly as for mtDNA Soqotra males have the lowest gene diversity from all neighboring populations under study.

Shared mtDNA haplotypes

The population of Soqotra has a large number of unique mitochondrial haplotypes (Table 2). Almost half of the Soqotri haplotypes (8/17) have never been detected in neighboring regions of southwestern Asia or Africa (compared to a database of 6,641 individuals living in neighboring areas). Moreover, two of the unique haplotypes are quite common on Soqotra; HVS-I motif 16086-16129-16148-16223 occurs at 23% and HVS-I motif 16126-16172-16355-16362 occurs at 15% in the Soqotri sample. The other six unique haplotypes are singletons with the exception of one haplotype that occurs in two individuals. Three of the singleton haplotypes are only a single mutational step away from the two high-frequency Soqotri haplotypes.

The largest number of matches with Soqotri haplotypes occurred in populations from the Arabian Peninsula (seven matches; 7.6% of Arabian haplotypes are shared with Soqotri haplotypes), Ethiopia (six matches; 6.3% of Ethiopian haplotypes are shared), and the Near East/Middle East/Northeast Africa (five matches each; 8.7, 6.0, and 7.0% of haplotypes are shared, respectively). The Near East and Middle East also had the highest frequency of a single haplotype match, i.e. the reference sequence (rCRS), which is the predominant haplotype in Europe, is absent in East Africa, and occurs at 6.3 and 4.5% in the Near East and Middle East, respectively. Only one or two haplotype matches were found in East Africa, Chad or India (Table 2).

Population comparisons

Based on F_{ST} values, the mitochondrial genetic diversity of Soqotra is statistically different ($P < 0.01$) from the comparative populations. An MDS plot of F_{ST} values shows that the Soqotra sample is clearly distinct from all sub-Saharan, North African, Middle East, and Indian populations (see Fig. 2). High differentiation of the East African groups such as the Sandawe, Hadza, Turu, Datog, and Burunge is shown on the left side of the graph. However, there is a general similarity of the remaining sub-Saharan African populations, particularly

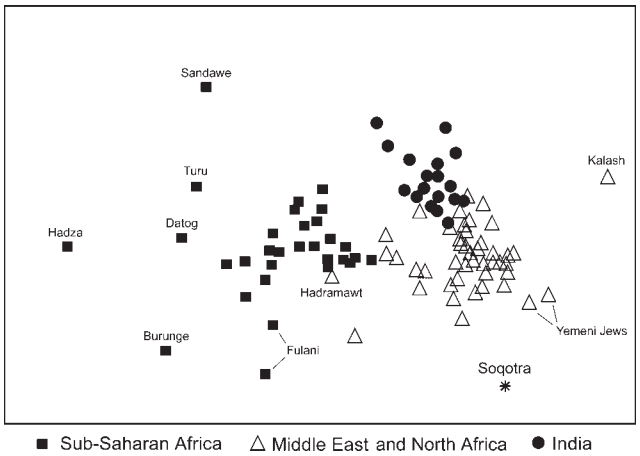

Fig. 2. MDS plot based on F_{ST} distances calculated from mtDNA sequences; Kruskal's stress value is 0.20936. Only the populations mentioned in the text are labeled.

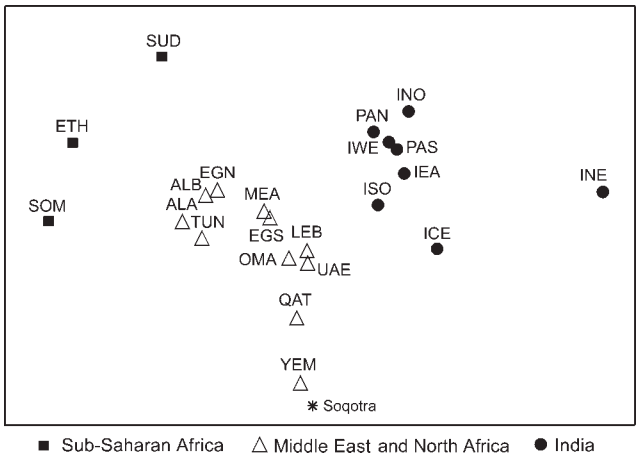

Fig. 3. MDS plot based on F_{ST} distances calculated from NRY haplogroup frequencies; Kruskal's stress value is 0.19938. For population codes, see Supporting Information Table 2.

those from the Sahel band and the Chad Basin (with the exception of the Fulani nomads). Subsequently, there is a transitional zone formed by the populations from Ethiopia and the Nile Valley but also by some Yemeni groups, particularly the ones from the eastern parts of the country (Hadramawt). Finally, the cluster on the right part of the graph is composed by the Indian populations on the top, the Near and Middle Eastern groups in the middle and the populations of the Arabian peninsula at the bottom; Yemeni Jews being slightly different. The only outlier within the region of southwestern Asia is the Kalash sample that is situated on the extreme right part of the graph (see also Quintana-Murci et al., 2004). There is a general cline among all populations in the MDS plot from the Soqotri population to a cluster of Middle East and North African populations that splits into sub-Saharan and Indian populations.

Population differentiation of Soqotra from African, Middle East and Indian populations based on NRY-SNP data manifests a similar picture although the comparative populations are different and fewer than in the mitochondrial DNA analysis (see Fig. 3). A comparison of F_{ST} values shows that the only population that is not significantly different from Soqotra is that from Yemen ($P > 0.01$). Similarly to mtDNA MDS plot, we observe a cline from the Soqotri population to a cluster of Middle East and North African populations that splits into sub-Saharan and Indian populations.

Phylogenetic affiliations

Within the Soqotri samples, we identified haplotypes belonging to three of the main branches of the mtDNA phylogeny (macrohaplogroups L, N, and R); notably haplogroup M is absent (Table 2). There are only two sub-Saharan L haplotypes and they do not carry the 3594*Hpa*I mutation so their classification is L3*; these haplotypes do not contain the specific mutations of L5b (−3594*Hpa*I) (Kivisild et al., 2004) and therefore they are possibly L3h2 as they both contain substitutions at 16111, 16184, and 16304 (see Behar et al., 2008). Macrohaplogroup N is represented by three different haplotypes of which only one can be unambiguously classified as N1a (it contains HVS-I motif 16147G-16172-16223-16248-16355). Two other N haplotypes have never been found outside Soqotra (see Table 2).

The most widespread mtDNA types in Soqotra belong to macrohaplogroup R (Table 2). The majority of R haplotypes can be classified as R0a [previously known as (preHV)1]. Three of the R haplotypes have not been previously reported. A network analysis of all Soqotri R0a haplotypes with additional sequences from Africa and Asia (see Fig. 4) shows a time to most recent common ancestor (TMRCA) of 23,339 ± 8,232 YBP for R0a. It is shown that the majority of Soqotri R0a haplotypes fall into clade R0a1 (defined by variant 16355) whose TMRCA is 11,418 ± 4,198 YBP. Furthermore, within R0a1, the unique Soqotri haplotypes form a new clade that is defined by variant 16172 and that we have named R0a1a1. Abu-Amero et al. (2007) identified a haplotype defined by variant 16355 and named it (preHV)1a1, thus it corresponds to R0a1a using the newer nomenclature and the unique Soqotri haplotypes are derived from this lineage). This Soqotri-specific clade has a very young TMRCA (3,363 ± 2,378 YBP) that suggests the R0a1a1 haplotypes evolved on Soqotra and have not dispersed elsewhere. Two other Soqotri R haplotypes are not classified further than R* and are quite common in neighboring populations. Five haplotypes within macrohaplogroup R carry the 4216*N1a*III variant that places them in clade JT. Of the JT haplotypes, two are unique to Soqotra; J1b is represented by two individuals and T* is represented by one individual.

The majority of NRY haplotypes in Soqotra belong to haplogroup J (85.7%), with most (45 out of 54) unclassified as J*(xJ1,J2) and a few (the remaining 9 samples) classified as J1 (see Fig. 5). It is interesting to note that NRY haplotypes lacking both M172 and M267, as in our unclassified J*, have not been previously identified on the Arabian Peninsula (Cadenas et al., 2008). Haplogroup E is represented at a frequency of 9.5% and three other haplogroups, F*(xJ,K), K*(xO,P) and R*(xR1b), are present in one individual each. It is worth noting that none of the ancient African haplogroups (A and B) were observed in Soqotra.

DISCUSSION

Our mitochondrial and Y chromosome analysis of Soqotri populations provides the first indication of how

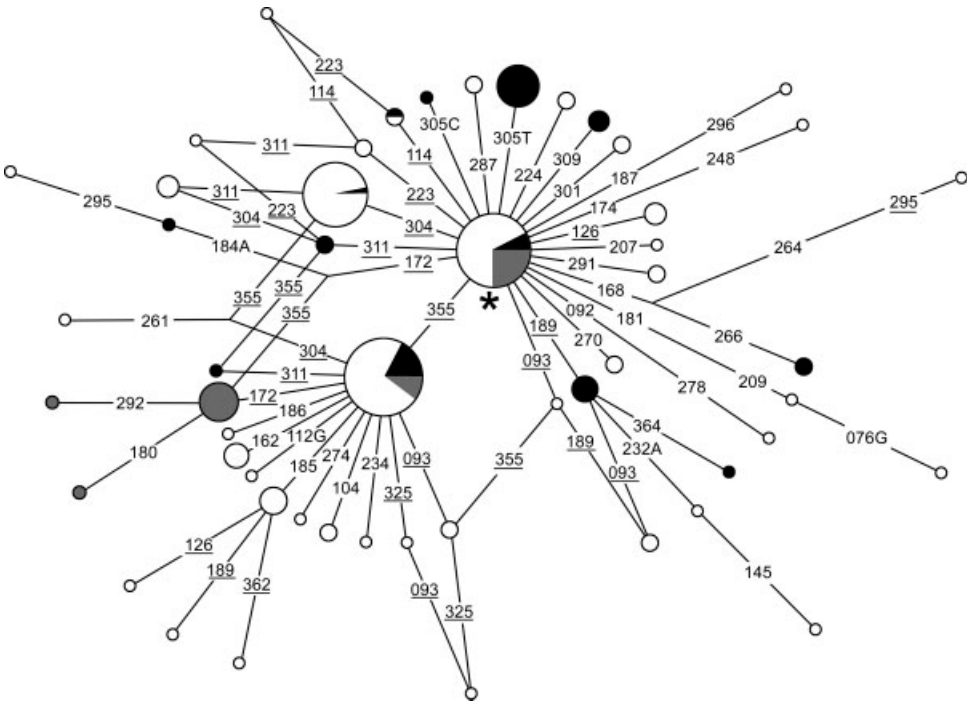

Fig. 4. Network of haplogroup R0a based on HVS-I sequences from Soqotra and neighboring regions (see Supp. Info. Table 1). Variant positions from 16,030 to 16,370 (minus 16,000) are shown. The R0a haplotype (16126–16362) is shown with an asterisk. Parallel mutations are underlined. Circle sizes are proportional to haplotype frequency. Black, white, and gray colors indicate samples from Africa, Asia, and Soqotra, respectively.

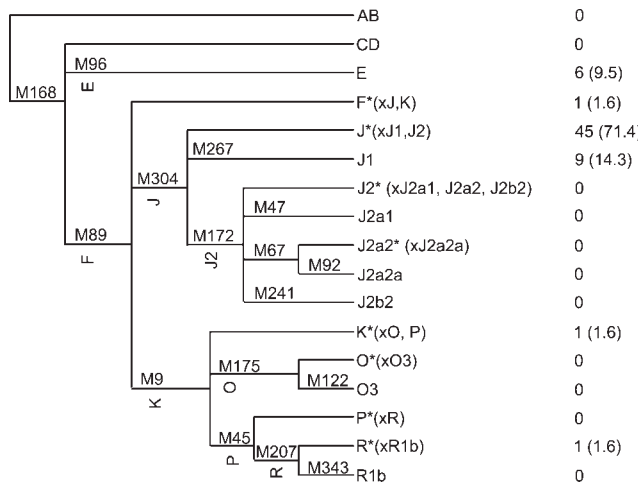

Fig. 5. Phylogeny of Y-chromosomal haplogroups; the names of haplogroups are given on the tip of the lineage according to Karafet et al., 2008. Polymorphisms screened in this study are shown along the branches. The absolute numbers of chromosomes and relative frequencies within each haplogroup are provided in the right column.

and when this island was originally colonized. Specifically, an analysis of mismatch distribution of mtDNA shows there is a signature of demographic expansion as Harpending's raggedness index does not attain significant value in Soqotra population sample (Table 1 for mtDNA). However, molecular diversity indices of both mtDNA and NRY show that the Soqotra population has rather reduced diversity compared to a subset of neighboring populations from Africa, Arabian Peninsula, Near East or India (Table 1) that is explicable as a result of a founder effect (see for example Cordaux et al., 2004; Abbott et al., 2006; Pichler et al., 2006).

As shown by the population comparisons, for both mtDNA and NRY, the Soqotra islanders show evidence of long-term isolation although relationships can be found with populations from the Arabian Peninsula. As evidenced from MDS, mtDNA variation shows less differentiation between regions than NRY variations; however, different molecular resolutions as well as different population samples used for mtDNA and NRY analyses should be taken into account. Notwithstanding, an Arabian origin of the Soqotri people is evidenced not only by MDS results but also by the highest number of mtDNA haplotypes shared with Arabia—7 out of 17 Soqotri haplotypes were reported in Arabia (see Table 2).

Mitochondrial and Y chromosome haplogroups in Socotra

The presence of mitochondrial macrohaplogroup L in Arabia has most often been explained as a result of the Arab slave trade or other recent gene flow (Richards et al., 2003; Kivisild et al., 2004; Abu-Amero et al., 2007). The highest frequency of L haplotypes in Arabia was detected in eastern Yemeni Hadramawt (60%), while the western part of Yemen contained a much lower frequency of sub-Saharan L haplotypes (Černý et al., 2008). As one moves further north on the Arabian Peninsula, the frequency of African haplotypes decreases. L haplotypes on the Arabian Peninsula can be differentiated as L3 haplotypes introduced from northern parts of East Africa, such as Ethiopia, Sudan and Egypt, and L0a types derived from southern parts of East Africa, such as Tanzania, Kenya, and Mozambique. Soqotra contains only two L-haplotypes that are classified as L3* (tentatively L3h2). One haplotype is unique and is two mutational steps away from the second haplotype, which has an exact match in Ethiopia (Table 2). Similar evidence of recent gene flow into Soqotra is provided by the Y chromosome

data presented here, in which no ancient African haplogroups (A, B) were detected. Although the subtypes of haplogroup E were not finely resolved in our analysis, the total contribution of E haplotypes to the Soqotra gene pool is not greater than 9.5%. This picture of genetic variation, therefore, does not support a high level of gene flow from Africa. Furthermore, the absence of mitochondrial macrohaplogroup M and haplogroups L5 and L6, suggests that Soqotra does not preserve evidence of the oldest human migration out-of-Africa that was estimated to have occurred ~65,000 years ago based on genetic evidence (Macaulay et al., 2005; Torroni et al., 2006).

Macrohaplogroup N (xR) is represented in Arabia at low frequencies and mainly by haplogroups N1a, N1b, N1c, I, W, and X2 (Kivisild et al., 2004; Abu-Amero et al., 2007; Rowold et al., 2007; Černý et al., 2008). In Soqotra, only N1a (highest number of matches in Ethiopia) and the unclassified N* (two unique lineages in Yemen) were identified. N1a seems to have been widespread in the Neolithic as it has been identified in Central Europe as well as in the Altai mountains (Ricaut et al., 2004; Haak et al., 2005). However, the highest frequency of this lineage today is in the Arabian Peninsula where it reaches the frequencies of 5% (Kivisild et al., 2004; Abu-Amero et al., 2007; Černý et al., 2008). Since both the highest diversity and ancestral haplotypes of N1a are present in the Arabian Peninsula (Abu-Amero et al., 2007), it seems that the lineage (and derived haplotypes with 16147A) dispersed from this region westwards to Europe and northwards to Central Asia and possibly also in a southward direction across the Gulf of Aden. Thus, the presence of N1a in Soqotra is not surprising. However, it is more difficult to interpret the presence of two unique N* haplotypes, at a combined frequency of 25%, in Soqotra. At a minimum, it seems that hpt03 (n = 1) evolved from hpt04 (n = 15), which is a founder lineage on Soqotra; there is insufficient variation to date this expansion, but the lack of variation suggests it must be very recent.

Macrohaplogroup R is represented by the haplogroups R0a, HV, H, V, U, J, and T in Western Asia but, in Soqotra we found only R0a, unclassified R*, J, and T, of which R0a is most prevalent at a frequency of 38%. Haplogroup R0a is represented by five different haplotypes, three of which have not been previously identified. Both R0a and R0a1 lineages [previously designated (preHV)1 and (preHV)1a, respectively, as in Abu-Amero et al., 2007] can be identified in the Soqotri dataset. We obtained slightly older TMRCA dates for R0a and R0a1 (23,339 ± 8,232 and 11,418 ± 4,198 YBP, respectively) than Abu-Amero et al. (2007; 18,993 ± 6,999 and 9,624 ± 2,994 YBP, respectively), although there is considerable overlap of each set of confidence intervals. The star-like subhaplogroup R0a1 shows a clear expansion during the Neolithic period as it is much more frequent in the Near East and Arabian Peninsula than elsewhere (the only African R0a1 sequences belong to the root haplotype). The unique Soqotri R0a1 haplotypes (designated R0a1a1 here) have a TMRCA of 3,363 ± 2,378 YBP and suggest a Holocene expansion of R0a1 on Soqotra between ~11,000 and ~3,000 years ago, i.e. the TMRCAs of R0a1 and R0a1a1, respectively.

The presence of unique mitochondrial haplotypes (N* and R0a1a1) in Soqotra is similar to the distribution of mitochondrial diversity in the Canary Islands off the northwest coast of Africa where there are several unique U6b sequences that seem to have undergone a recent expansion throughout the Canary Islands (Rando et al., 1999). Rando et al. (1999) proposed a total of six founding lineages for the Canarian settlement and pooled the variation across all six haplogroups to estimate an age of Canary-specific variation of 2,800 ± 900 years old, concordant with archaeological evidence. If we also pool the variation at the two most evident founder lineages in Soqotra (N* and R0a1a1; hpts4 and 10 in Table 2), there are three mutational events in 28 individuals, which corresponds to an age of 2,162 ± 1,248 YBP. If we also include L3* (hpt 2), which is not as clearly a founder lineage, there are five mutational events in 31 individuals, equivalent to 3,255 ± 1,456 YBP. These estimates are consistent with a late Holocene settlement of Soqotra.

For the perspective of the Y-chromosome data, a high frequency of haplogroup J1 in Soqotra is consistent with a gradient of this haplogroup in the Arabian Peninsula (Cadenas et al., 2008). These authors estimated ages for J1 in Arabia (9.7 ± 2.4 in Yemen, 7.4 ± 2.3 in Qatar and 6.4 ± 1.4 KYBP in UAE), consistent with a Neolithic expansion from the north (where Y-chromosome STR diversity is higher). However, we report a much higher frequency of J* (lack of M267 and M172) in Soqotra. Since this lineage was not found by Cadenas et al. (2008) in the Arabian Peninsula, this raises the possibility of an earlier input for these lineages or more probably very strong genetic drift of a low frequency Arabian lineage in the Y-chromosome gene pool of Soqotra.

CONCLUSIONS

We present the first molecular genetic study of human populations living on the main island of the Soqotra archipelago. We found that almost half of the Soqotri mitochondrial haplotypes (8/17) have never been detected in neighboring regions of southwestern Asia or Africa. Further, Soqotri genetic variation is characterized by at least two autochthonous mitochondrial clades that attest to the long-term isolation of the island. Expansion of these clades seems very recent, possibly as recent as ~3,000 years ago, supporting a late Holocene settlement of the island. On Soqotra, the presence of African and Eurasian lineages at frequencies most similar to Arabian Peninsula populations and the sharing of haplotypes with Arabian populations is consistent with a major, recent colonization of Soqotra from southern Arabia.

ACKNOWLEDGMENTS

The authors express gratitude to the Environment Protection Authority, Ministry of Water and Environment, Republic of Yemen for providing the research permit for Soqotra. We are highly indebted to Chris Edens for his support at the American Institute for Yemeni Studies. We thank the anonymous reviewers for valuable comments. Last, but not least, we would like to express our gratitude to the Soqotri volunteers for their participation in this study.

LITERATURE CITED

Abu-Amero KK, Gonzalez AM, Larruga JM, Bosley TM, Cabrera VM. 2007. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol 7:32.

Abbott WG, Winship IM, Gane EJ, Finau SA, Munn SR, Tukuitonga CE. 2006. Genetic diversity and linkage disequilibrium in the Polynesian population of Niue Island. Hum Biol 78: 131–145.

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt HJ, Semino O, Torroni A. 2005. Saami and Berbers–an unexpected mitochondrial DNA link. Am J Hum Genet 76:883–886.

Bandelt H-J, Macaulay V, Richards M. 2006. Human mitochondrial DNA and the evolution of Homo sapiens. Berlin, New York: Springer.

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. Genetics 141:743–753.

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S. 2008. The dawn of human matrilineal diversity. Am J Hum Genet 82:1130–1140.

Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ. 2008. Y-chromosome diversity characterizes the Gulf of Oman. Eur J Hum Genet 16:374–386.

Černý V, Hájek M, Bromová M, Čmejla R, Diallo I, Brdička R. 2006. MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. Hum Biol 78:9–27.

Černý V, Hájek M, Čmejla R, Brůžek J, Brdička R. 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. Ann Hum Biol 31:554–569.

Černý V, Mulligan CJ, Rídl J, Žaloudková M, Edens CM, Hájek M, Pereira L. 2008. Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. Am J Phys Anthropol 136:128–137.

Cheung C, DeVantier L. 2006. Socotra—a natural history of the islands and their people (Odyssey Books and Guides). UK: NHBS Environment Bookstore.

Cordaux R, Weiss G, Saha N, Stoneking M. 2004. The northeast Indian passageway: a barrier or corridor for human migrations? Mol Biol Evol 21:1525–1533.

Excoffier LGL, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.

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

Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tanzer M, Villems R, Renfrew C, Gronenborn D, Alt KW, Burger J. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310:1016–1018.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98.

Harpending HC. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum Biol 66:591–600.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75:752–770.

Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Paabo 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.

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

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5:26.

Morris M. 2002. Manual of traditional land use in the Soqotra archipelago. G.E.F. report no. YEM/96/G32. Royal Botanic Garden Edinburgh.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

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

Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75:966–978.

Pichler I, Mueller JC, Stefanov SA, De Grandi A, Volpato CB, Pinggera GK, Mayr A, Ogriseg M, Ploner F, Meitinger T, Pramstaller PP. 2006. Genetic structure in contemporary south Tyrolean isolated populations revealed by analysis of Y-chromosome, mtDNA, and Alu polymorphisms. Hum Biol 78:441–464.

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, Qasim Mehdi S, 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.

Rando JC, Cabrera VM, Larruga JM, Hernandez M, Gonzalez AM, Pinto F, Bandelt HJ. 1999. Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. Ann Hum Genet 63:413–428.

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk HV, Parik J, Loogvali EL, Derenko M, Malyarchuk B, Bermisheva M, Zhadanov S, Pennarun E, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechanina E, Mikerezi I, Moisan JP, Chaventre A, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Rickards O, De Stefano GF, Papiha S, Beckman L, Janicijevic B, Rudan P, Anagnou N, Michalodimitrakis E, Koziel S, Usanga E, Geberhiwot T, Herrnstadt C, Howell N, Torroni A, Villems R. 2003. Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet 73:1178–1190.

Ricaut FX, Keyser-Tracqui C, Bourgeois J, Crubezy E, Ludes B. 2004. Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient central Asian migrations. Hum Biol 76:109–125.

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.

Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, Villems R. 2007. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. Mol Biol Evol 24:436–448.

Rowold DJ, Luis JR, Terreros MC, Herrera RJ. 2007. Mitochondrial DNA geneflow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. J Hum Genet 52:436–447.

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby 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 MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111.

Serjeant RB. 1963. The Portuguese off the South Arabian coast; Hadrami Chronicles, with Yemeni and European accounts of Dutch pirates off Mocha in the seventeenth century. Oxford: Clarendon Press.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

Tajima F. 1993. Measurement of DNA polymorphism. In: Takahata N, Clark AG, editors. Mechanisms of molecular evolution introduction to molecular paleopopulation biology. Tokyo, Sunderland, MA: Japan Scientific Societies Press, Sinauer Associates. p 37–59.

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L. 2005. Reconstructing the origin of Andaman Islanders. Science 308:996.

Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339–345.