

ARTICLE

Y-chromosome diversity characterizes the Gulf of Oman

Alicia M Cadenas¹, Lev A Zhivotovsky², Luca L Cavalli-Sforza³, Peter A Underhill³ and Rene J Herrera^{*1}

¹Department of Biological Sciences, Florida International University, Miami, FL, USA; ²N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia; ³Department of Genetics, Stanford University, Stanford, CA, USA

Arabia has served as a strategic crossroads for human disseminations, providing a natural connection between the distant populations of China and India in the east to the western civilizations along the Mediterranean. To explore this region's critical role in the migratory episodes leaving Africa to Eurasia and back, high-resolution Y-chromosome analysis of males from the United Arab Emirates (164), Qatar (72) and Yemen (62) was performed. The role of the Levant in the Neolithic dispersal of the E3b1-M35 sublineages is supported by the data, and the distribution and STR-based analyses of J1-M267 representatives points to their spread from the north, most likely during the Neolithic. With the exception of Yemen, southern Arabia, South Iran and South Pakistan display high diversity in their Y-haplogroup substructure possibly a result of gene flow along the coastal crescent-shaped corridor of the Gulf of Oman facilitating human dispersals. Elevated rates of consanguinity may have had an impact in Yemen and Qatar, which experience significant heterozygote deficiencies at various hypervariable autosomal STR loci. *European Journal of Human Genetics* (2008) 16, 374–386; doi:10.1038/sj.ejhg.5201934; published online 10 October 2007

Keywords: Arabia; Y-chromosome; SNP; Y-STR

Introduction

Arabia has played the role of a strategic crossroads between Africa and Eurasia, facilitating the first exodus of modern humans from the Horn of Africa to the present day Yemen through the Bab el Mandab Strait at mouth of the Red Sea.^{1–3} Subsequent migrations through the northern intercontinental passageway between Africa and the Levant (the Levantine corridor) have also been documented.^{4,5} In addition, the Arabian Peninsula has linked the distant populations of China and India to communities of the Mediterranean and beyond. Although the Persian Gulf

to the east and the Arabian Sea to the south offered easy passages to India and Asia, the Red Sea on the western coast of the Arabian Peninsula provided a natural connection to the Mediterranean Sea.

Just north of the peninsula, the Nile River Valley in Egypt and the Tigris–Euphrates area in Iraq comprised a region known as the Fertile Crescent. Recognized as the birthplace of agriculture during the Neolithic (~8000 yBP) based on linguistic and archaeological evidence,^{6,7} the Fertile Crescent participated in ancient international trade. Although the fertile soils produced a surplus of food, the region lacked the natural resources necessary for building permanent structures (timber) or making metals (minerals). Therefore, early inhabitants relied on trade to acquire these raw materials and established close links with the commercial centers along the Persian Gulf as reflected in archaeological finds.^{8–10} At the extreme southern end of the Arabian Peninsula, referred to as Arabia Felix by the Romans

*Correspondence: Dr RJ Herrera, Department of Biological Sciences, Florida International University, University Park, OE 304, Miami, FL 33199 USA.

Tel: +1 305 348 1258; Fax: +1 305 348 1259;

E-mail: herrerar@fiu.edu

Received 20 February 2007; revised 30 August 2007; accepted 11 September 2007; published online 10 October 2007

(‘Happy Arabia’ in Latin) and including present day Yemen, the spice trade was an important source of wealth. Frankincense and myrrh were commonly exported to the Mediterranean via camels and to India by sea.

In agreement with archaeological and historical records that accentuate the region’s active role as a point of contact between distant populations, the Middle East displays a high degree of genetic diversity.^{11–14} Although genetic diversity is elevated, various analyses have identified structural barriers to gene flow into and out of the Near East. Specifically, mtDNA,¹⁵ Y-chromosome^{14,16–18} and autosomal STR studies¹⁹ have identified the Dasht-e Kavir and Dash-e Lut deserts in Iran and the Hindu Kush mountains in eastern Afghanistan as potential barriers to gene flow to the surrounding regions. In contrast, geographic facilitators for gene flow have also been described, including a region along the southern coast of Iran, Afghanistan and Pakistan known as Balochistan mediating gene flow from South Pakistan to South Iran.¹⁴

Mitochondrial DNA analyses have been performed on collections from Qatar, United Arab Emirates (UAE) and Yemen,^{5,20} yet the paternal component of this historically and geographically significant region is incomplete. Although Y-chromosome studies have focused on neighboring areas, including Egypt,⁴ Somalia,²¹ Iraq,²² Syria and Lebanon²³ as well as on the southern Arabian populations of Oman⁴ and Yemen,²⁴ high resolution Y-chromosome analyses of the Persian and Oman Gulfs are fragmentary.

To gain a more complete understanding of this region’s role in human dispersals, particularly in light of previous studies that have identified barriers and conduits to gene flow that would affect its Y-haplogroup substructure, the present study employs high-resolution Y-chromosome analyses of three southern Arabian populations: Yemen ($n = 62$), Qatar ($n = 72$) and the UAE ($n = 164$). In addition, 17 Y-STR loci were typed to obtain STR-based age estimates for a selection of informative Y-chromosome haplogroups in the populations in which they were observed. Results from these Y-specific analyses were interpreted in conjunction with data on 15 autosomal STR loci for Yemen, Oman, Qatar, Iran, Egypt^{19,25,26} and UAE (Cadenas, unpublished results) reanalyzed collectively with the aim of exposing characteristics unique to the southern Arabian Peninsula.

Materials and methods

Sample collection and DNA extraction

Blood samples from 298 unrelated males representing three populations that include the UAE, Qatar and Yemen were collected in EDTA Vacutainer tubes. The paternal ancestry of the donors was recorded for a minimum of two generations. Table 1 provides additional information on the sample size, geography and linguistic affiliation of the populations involved. DNA was extracted from the blood using the phenol–chloroform extraction method.³³ Ethical

guidelines were adhered to in strict compliance with NIH guidelines as well as to those stipulated by the institutions involved.

Y-haplogroup analysis

Seventy-six binary genetic markers were genotyped^{12,30,34–36} following the Y-chromosome phylogeny hierarchy using standard methods, including PCR/RFLP, allele-specific PCR³⁷ and the YAP polymorphic *Alu* insertion.³⁸ The amplicons generated from these methods were separated by electrophoresis in 1X TAE, 3% agarose gels and visualized subsequent to ethidium bromide staining and UV light photography in a Fotodyne FOTO/Analyst[®]. The phylogenetic relationships of the relevant Y-chromosome haplogroups are illustrated in Figure 1 according to YCC nomenclature³⁹ with new marker designations as provided in the published literature.^{12,30,34–36}

Statistical and phylogenetic analyses

Twenty-nine geographically targeted populations reported in previous studies (Table 1) were included in the statistical and phylogenetic analyses performed to assess Y-haplogroup variation and phylogeographic relationships throughout the region. The Georgia and Tajikistan data will be published in detail elsewhere. The various data sets were used at a resolution of major haplogroups (A through R). Haplogroup frequencies were compared by means of a χ^2 -test. Phylogenetic comparisons were made with multidimensional scaling (MDS) analysis based on *Fst* distances⁴⁰ using the Statistical Package for the Social Sciences (SPSS) software program.⁴¹ Genetic structure was further examined by performing two sets of analyses of molecular variance (AMOVA)⁴² using the Arlequin version 2.000 package⁴³ with the 32 populations subdivided according to two criteria, geography (North Africa, East Africa, Arabian Peninsula, Caucasus, Levant, Anatolia, Iranian Plateau, South Asia and Central Asia) and linguistic family (Afro-Asiatic, Indo-European, Niger-Congo, Altaic and South Caucasian). Table 1 indicates the populations included in each of the geographic and linguistic groups utilized in the AMOVA. Pairwise comparisons of the populations from the present study and all reference populations were generated using *G*-tests in Carmody’s software⁴⁴ to assess any genetic differences of statistical significance.

STR analysis

DNA amplification of 17 Y-specific STR loci (specifically DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and Y-GATA H4) was performed using the AmpFISTR Yfiler Amplification Kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer’s instructions in an Eppendorf[®] Mastercycler[®]. DNA fragment separation and detection was achieved in an ABI

Table 1 Geographic and linguistic description of populations analyzed

Geographic region	Population	Abbreviation	n	Language family	Reference
North Africa	Egypt	EGY	147	Afro-Asiatic	Luis <i>et al</i> ⁴
	Algeria (Arabs)	ALA	35	Afro-Asiatic	Arredi <i>et al</i> ²⁷
	Algeria (Berbers)	ALB	19	Afro-Asiatic	Arredi <i>et al</i> ²⁷
	Tunisia	TUN	148	Afro-Asiatic	Arredi <i>et al</i> ²⁷
East Africa	Kenya	KEN	29	Niger-Congo	Luis <i>et al</i> ⁴
	Somalia	SOM	201	Afro-Asiatic	Sanchez <i>et al</i> ²¹
	Oromo	ORO	78	Afro-Asiatic	Semino <i>et al</i> ²⁸
	Amhara	AMH	48	Afro-Asiatic	Semino <i>et al</i> ²⁸
Arabian Peninsula	United Arab Emirates	UAE	164	Afro-Asiatic	Present study
	Qatar	QAT	72	Afro-Asiatic	Present study
	Yemen	YEM	62	Afro-Asiatic	Present study
	Oman	OMA	121	Afro-Asiatic	Luis <i>et al</i> ⁴
Caucasus	Georgia	GEO	48	South Caucasian	Regueiro (unpublished results)
	Armenia	ARM	100	Indo-European	Nasidze <i>et al</i> ²⁹
	Azerbaijan	AZE	72	Altaic	Nasidze <i>et al</i> ²⁹
Levant	Lebanon	LEB	31	Afro-Asiatic	Semino <i>et al</i> ²³
	Syria	SYR	20	Afro-Asiatic	Semino <i>et al</i> ²³
Anatolia	Iraq	IRQ	139	Afro-Asiatic	Al-Zahery <i>et al</i> ²²
	Greece	GRE	76	Indo-European	Semino <i>et al</i> ²³
	Turkey	TUR	523	Altaic	Cinnioğlu <i>et al</i> ¹²
Iranian Plateau	North Iran	NIR	33	Indo-European	Regueiro <i>et al</i> ¹⁴
	South Iran	SIR	117	Indo-European	Regueiro <i>et al</i> ¹⁴
South Asia	North Pakistan	NPK	85	Indo-European	Sengupta <i>et al</i> ³⁰
	South Pakistan	SPK	91	Indo-European	Sengupta <i>et al</i> ³⁰
	Gujarat	GUJ	29	Indo-European	Kivisild <i>et al</i> ³¹
	Punjab	PUN	66	Indo-European	Kivisild <i>et al</i> ³¹
	Mongolia	MON	147	Altaic	Karafet <i>et al</i> ³²
Central Asia	Tajikistan	TAJ	24	Indo-European	Regueiro (unpublished results)
	Uzbekistan	UZB	54	Altaic	Karafet <i>et al</i> ³²
	Kazakhstan	KAZ	30	Altaic	Karafet <i>et al</i> ³²
	Kyrgyzstan	KYR	52	Altaic	Wells <i>et al</i> ¹⁷
	Turkmenistan	TKM	30	Altaic	Wells <i>et al</i> ¹⁷

Prism 3100 Genetic Analyzer (Applied Biosystems). ABI Genescan500 LIZ was utilized as an internal size standard. Amplicon sizes were determined using the Genescan[®] 3.7 software and alleles were designated by comparison to an allelic ladder from the manufacturer using Genotyper[®] 3.7 NT software.

Haplogroup-specific expansion times were estimated for select binary haplogroups (J1-M267, R1a1-M198, E3b1a-M78 and E3b1c-M123) by the linear expansion method. This procedure assumes a stepwise mutation model⁴⁵ and a mean STR mutation rate of 0.00069 per STR locus per generation⁴⁶ with a 25-year intergeneration time as performed in previous studies.^{4,12,47} The linear expansion method assumes a star-like genealogy attributable to continuous growth where the expected value of the average coalescence time for STR alleles (T) equals the STR variance (S) divided by the mutation rate (μ) times the number of generations since expansion.^{48,49} STR variances were calculated using the vp equation of Kayser *et al*.⁵⁰ In addition, STR-based divergence times were calculated for each of the haplogroups based on the method described by Zhivotovsky *et al*,^{46,51} likewise assuming a mutation rate of 0.00069 per STR locus per generation⁴⁶ and a 25 year intergeneration time.

Autosomal STR markers

To further assess the level of homogeneity of the populations under study, the observed and expected heterozygosity for 15 autosomal STR loci were calculated using the Arlequin version 2.000 package⁴³ based on the genotypes for Iran,²⁵ Qatar,²⁶ UAE (Cadenas, unpublished results), and Kenya, Egypt, Oman and Yemen.¹⁹ Heterozygote deficiencies (Fis and corresponding P -values) were computed for all seven populations using the GENEPOP2 software⁵² according to the test described by Rousset and Raymond.⁵³

Results

Phylogeography

A total of 41 paternal haplogroups were identified from the analysis of 164 United Arab Emirate, 72 Qatari and 62 Yemeni males. Figure 1 displays their hierarchical phylogeny as well as the frequency (percentages) distributions for the populations under study. The geographic distribution of the major haplogroups is illustrated in Figure 2. Only three haplogroups (E, J and R) display frequencies above 5% in the three populations occupying the southern portion of the Arabian Peninsula and combined account

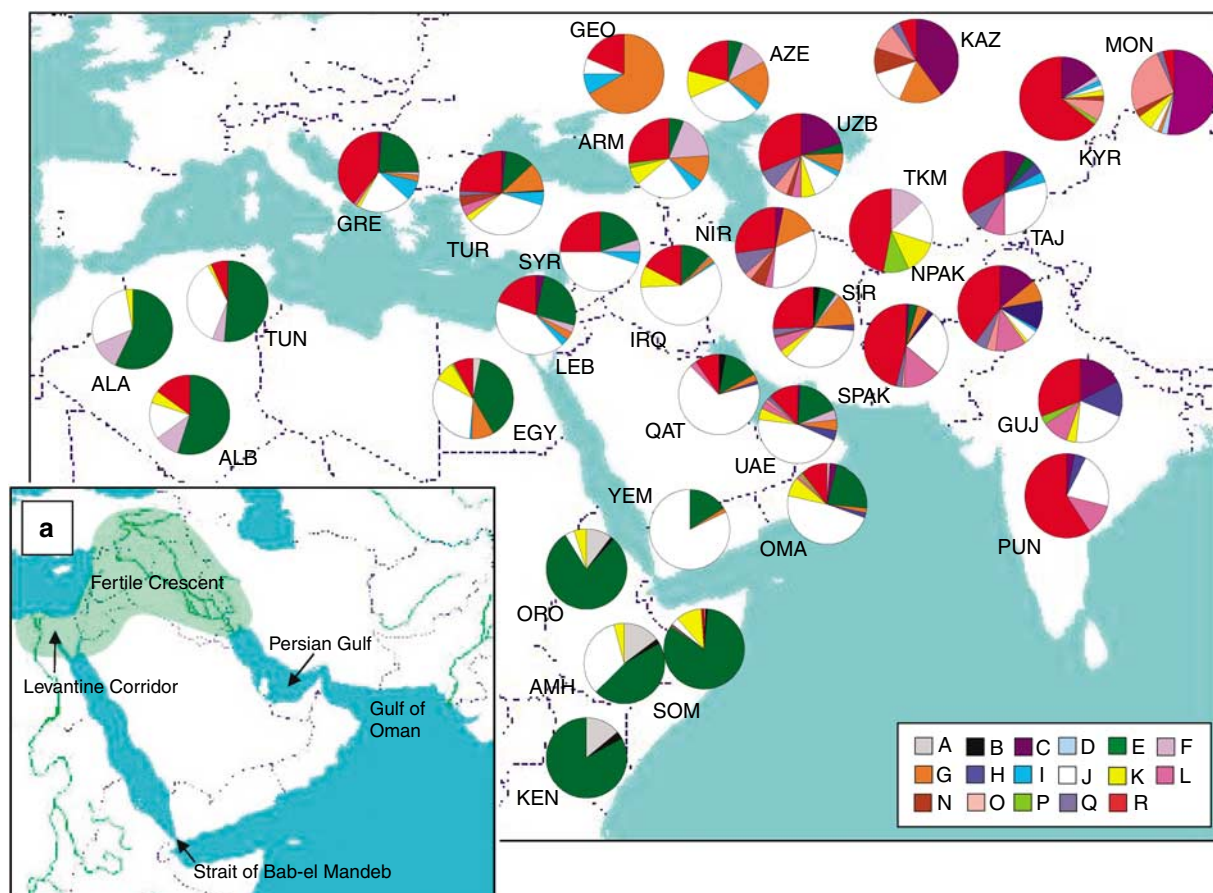


Figure 2 Geographic distribution of major Y-chromosome haplogroup frequencies for the 32 populations described in Table 1. (a) Geographic location of regions described in the text.

for 74–98% of the chromosomes within these collections. Figure 3 displays the geographic distribution of E, J and R derivatives in a subset of the populations listed in Table 1.

AMOVA

To explore potential correlations between genetic diversities and linguistic or geographic partitioning, the AMOVA was performed. The results of the AMOVA for the three populations from the present study in addition to 29 reference populations are listed in Table 2. Assignment of the populations according to the nine geographical groups described in Table 1 generated a higher fraction of variability among groups of populations (15.73%) than among populations within groups (4.21%) indicating a greater degree of interregional structuring. In contrast, upon subdividing the populations according to the language family (Table 1), the percentage of among group variance is lower (8.17%) than among populations within groups (12.94%), suggesting higher intralinguistic structuring for these populations. Genetic diversity among groups of

populations and populations within groups correlate significantly with geographic and linguistic partitioning.

Phylogenetic analyses

An MDS test was performed to assess phylogenetic relationships among populations. The MDS analysis performed on a matrix of F_{st} values based on haplogroup frequencies for the populations in Table 1 is displayed in Figure 4. Geographic structuring is observed involving populations displaying affiliations with other populations within their biogeographic zone. Within the plot, the populations of Egypt, Iraq, Yemen, Qatar, Oman, UAE, Syria and Lebanon occupy an intermediate position with populations from Africa to one side and Anatolia, Caucasus, Iranian Plateau, Central Asia and South Asia on the other. Of note, for the observed partitioning is the affinity of Egypt to populations from the Arabian Peninsula. Furthermore, Yemen and Qatar segregate together but separate from their neighboring populations, Oman and UAE particularly along Dimension 2. As expected, the populations from Central Asia group together and away

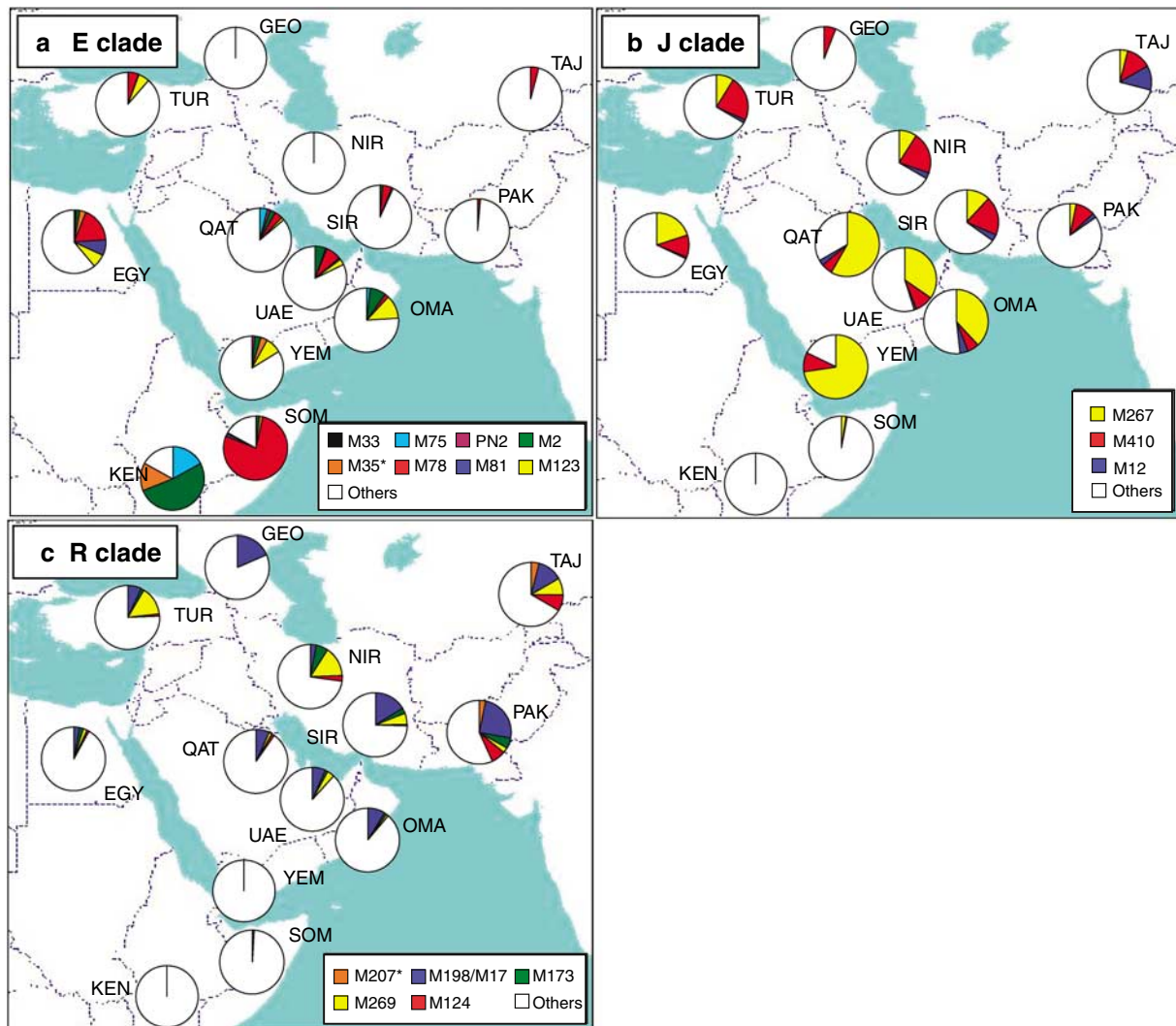


Figure 3 Geographic distribution of Y-chromosome haplogroup frequencies in selected populations (a) E clade, (b) J clade and (c) R clade.

Table 2 AMOVA results^a

	Variance (%)	Φ_{ST}	Φ_{SC}	Φ_{CT}
<i>Geography</i>				
Among groups	15.73			0.16**
Among populations within groups	4.21		0.05**	
Within population	80.06	0.20**		
<i>Language (family)</i>				
Among groups	8.17			0.08**
Among populations within groups	12.94		0.14**	
Within population	78.89	0.21**		

Φ_{ST} , variance within populations; Φ_{SC} , variance among populations within groups; Φ_{CT} , variance among groups.

** $P < 0.001$.

^aAMOVA was performed according to the geographic regions and language families as indicated in Table 1.

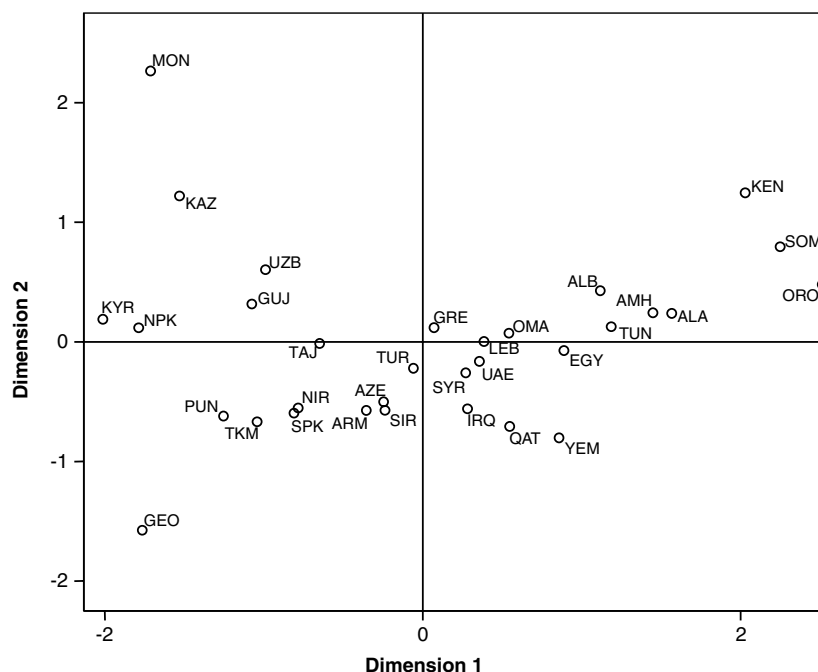


Figure 4 MDS analyses using F_{st} distances based on Y-haplogroup frequency data from the 32 populations described in Table 1.

from the South Asian ones, however, there is a segregation of North Pakistan with populations from Central Asia, whereas South Pakistan shares a closer affinity to populations to the west.

Interpopulation haplogroup diversity

To discern statistically significant genetic differences, pairwise G-test comparisons were performed with the three populations from the present study as well as the 29 reference populations described in Table 1. A total of 496 pairwise assessments were made and the results are provided as Supplementary Table 1. The number of nonsignificant genetic differences observed was 31. Of note, the Lebanese and Syrian populations do not display statistical differences with populations from UAE, Qatar and Oman but do show significant differences with Yemen. Furthermore, Yemen is the only one to display statistically significant differences to all other populations in the analysis. Across all populations, the Lebanese and Syrians are involved in most of the pairings in which no significant difference is observed ($\alpha = 0.05$). In addition to the three south Arabian populations of Qatar, UAE and Oman, the Algerian Berbers, Greece and Tajikistan do not exhibit significant differences to Lebanon and Syria. However, Turkey and South Iran only exhibit nonsignificant values in pairwise comparisons with Syria.

Y-STR diversity

Y-chromosome STR diversity was ascertained to generate haplogroup age estimates. Results from the expansion time

analysis of populations for haplogroups J1-M267, R1a1-M198, E3b1a-M78 and E3b1c-M123 are provided in Table 3. Y-STR data for the individuals genotyped are provided in Supplementary Table 2 (J1-M267), 3 (R1a1-M198), 4 (E3b1a-M78) and 5 (E3b1c-M123). The STR-based divergence times obtained using the method described by Zhivotovsky *et al*⁴⁶ (Table 3) for M123 are comparable (11.1 ky for UAE and 10.6 ky for Yemen), however, the UAE and Yemeni haplotypes within this haplogroup are quite different from each other and did not form a compact network, suggesting the ancestors for the M123 chromosomes in both populations involved independent bottlenecks followed by similar demographic processes. In contrast, the J1-M267 haplotypes formed a compact network across all the three populations and generated older age estimates for Yemen, Qatar and UAE (9.7, 7.4 and 6.4 ky, respectively) in comparison to the linear expansion method (7.11, 4.93 and 5.43 ky, respectively).

Intrapopulation autosomal STR diversity

In order to determine the level of heterozygote deficiency, possibly resulting from consanguinity, autosomal STR diversity was examined. Table 4 presents the observed and expected heterozygosity values of 15 autosomal STR loci for Kenya, Egypt, Iran, UAE, Oman, Yemen and Qatar with the F_{is} values provided in the final column for each population. A gradient in the number of loci that exhibit significant ($P < 0.05$) heterozygote deficiency is apparent moving north then east with lower values from Africa (1/15 loci for Kenya and Egypt) and higher amounts to the

Table 3 Y-Haplogroup variance, expansion and coalescence times based on Y-microsatellite loci

Population and Haplogroup	N	Haplogroup variance	T ^c (ky) ^d	Divergence time ^a (ky) ^b Mean ± SE
Qatar				
J1-M267	41	0.14	4.93	7.4 ± 2.3
R1a1-M198	5	0.19	6.71	14.5 ± 5.9
UAE				
E3b1a-M78	13	0.14	5.15	4.6 ± 1.8
E3b1c-M123	5	0.25	8.93	11.1 ± 3.9
J1-M267	57	0.15	5.43	6.4 ± 1.4
R1a1-M198	11	0.31	11.11	11.0 ± 3.3
Yemen				
E3b1c-M123	5	0.14	4.93	10.6 ± 4.1
J1-M267	44	0.20	7.11	9.7 ± 2.4

^aDivergence time based on SNP-STR coalescence method (Zhitovitsky^{46,51}).

^bBased on 15 Y-STR loci.

^cLinear expansion time assuming continuous growth and a 25-year intergeneration time.

^dBased on 8 Y-STR loci in common with previous studies (Cinniöglu *et al*,¹²; Luis *et al*,⁴).

west and south from the Iranian Plateau (2/15 in Iran) toward the southern populations of the Arabian Peninsula (2/15 for UAE, 3/15 for Oman, 4/15 for Yemen and 8/15 for Qatar).

Discussion

Analyses of the South Arabian Y-haplogroup substructure as well as the region's phylogenetic relationships to neighboring populations have provided us information on the following points: (1) support of the role of the Levant in the Neolithic dispersal of the E3b1-M35 derivatives, (2) neolithic spread of the J1-M267 haplogroup from the north, (3) a high haplogroup diversity shared among populations along the eastern and western coasts of the Gulf of Oman and (4) a limited haplogroup diversity in Yemen also supported by significant heterozygote deficiencies at various hypervariable autosomal STR loci.

Distribution of E3b1-M35 derivatives

The presence of signature sub-Saharan African mtDNA lineages in the south Arabian populations has been attributed to various waves of gene flow to the region, including that associated with the East African slave trade. This is apparent from the exact mtDNA haplotype matches between lineages in Yemen and East Africa, including those associated with the Bantu expansion.²⁰ The presence of the E3a-M2 lineage in Oman (7.4%),⁴ Yemen (3.2%), UAE (5.5%) and Qatar (2.8%) could lead to the oversimplified conclusion that these chromosomes are also a contribution from the East African slave trade. Mitochondrial DNA analysis of the Yemen Hadramawt indicates recent gene flow (~2500 yBP) from Africa to the Arab populations in part through the slave trade, yet an ancient arrival from East Africa is responsible for the Y-chromosome haplo-

types.⁵⁴ The contrast between female- versus male-mediated gene flow between these two areas can be attributed to the assimilation of females within the Arabian populations, whereas the males were often excluded from reproductive opportunities. The E3b1-M35 sub-haplogroups, M123 and M78, are believed to have spread from East Africa to North Africa and later expanded eastward through the Levantine corridor and westward to north-western Africa. Although E3b1a-M78 data suggest that this dispersal occurred in both directions,^{4,34,47} E3b1c-M123 disseminated primarily to the east.⁴ The distribution of the E3b1-M35 derivatives in Yemen, Qatar and UAE agrees with their arrival by expansion via the Levantine corridor rather than through the Horn of Africa. This route is similar to general patterns of Levantine mtDNA gene flows during the Upper Paleolithic⁵⁵ to the Neolithic.^{5,55} This is immediately apparent by the M35 profile of several East African populations and persisting even after introduction of E3a-M2 during the Bantu expansion, E3b1*-M35 is completely absent from the Omani,⁴ Qatari and UAE collections and relatively low in the Yemeni (3.2%). Kenya, Sudan and Tanzania^{4,56,57} also lack the E3b1c-M123 derivative that is common in the Near East.^{12,56–58} Furthermore, Ethiopia⁵⁶ and Somalia²¹ exhibit high levels of E3b1a-M78 (22.7 and 77.6%, respectively), which is null or nearly absent in the two populations closest to the Strait of Sorrows (Bab-el Mandeb Channel), Yemen (0%) and Oman (1.7%),⁴ ($\chi^2 = 170.618$, d.f. = 1, $P < 0.0001$ when combining the frequencies for Ethiopia and Somalia versus Yemen and Oman).

On the other hand, Cruciani *et al*⁵⁷ have postulated that the E3b1c-M123 clade may have originated in the Near East, as its presence in East Africa is restricted to Ethiopia (11.2%). The median expansion time for M123 in Egypt is

Table 4 Heterozygote deficiency based on 15 autosomal STR loci

Locus	Kenya			Egypt			Iran			UAE			Oman			Yemen			Qatar			
	N	He	Fis	N	He	Fis	N	He	Fis	N	He	Fis	N	He	Fis	N	He	Fis	N	He	Fis	
D8S1179	65	0.778	0.831	119	0.832	0.815	150	0.827	0.860	52	0.833	0.788	141	0.841	0.844	59	0.835	0.864	133	0.842	0.820	0.026
D21S11	65	0.851	0.862	119	0.803	0.815	150	0.851	0.767	52	0.852	0.846	141	0.854	0.851	59	0.825	0.881	132	0.811	0.811	0.001
D7S820	65	0.760	0.769	119	0.775	0.815	150	0.790	0.720	52	0.797	0.769	141	0.775	0.730	59	0.776	0.746	130	0.797	0.769	0.035
CSF1PO	65	0.766	0.785	119	0.723	0.731	150	0.713	0.693	52	0.738	0.692	141	0.730	0.709	59	0.719	0.763	131	0.728	0.710	0.025
D3S1358	65	0.749	0.785	119	0.747	0.723	150	0.766	0.720	52	0.758	0.712	141	0.785	0.745	59	0.738	0.712	133	0.785	0.744	0.052
TH01	65	0.764	0.800	119	0.754	0.773	150	0.788	0.793	52	0.748	0.712	141	0.791	0.716	59	0.769	0.729	133	0.760	0.699	0.080*
D13S317	65	0.742	0.785	119	0.783	0.706	150	0.763	0.780	52	0.807	0.788	141	0.779	0.716	59	0.753	0.793	131	0.780	0.618	0.208*
D16S539	65	0.793	0.723	119	0.781	0.714	150	0.774	0.713	52	0.766	0.731	141	0.779	0.752	59	0.781	0.797	132	0.749	0.674	0.100*
D16S539	65	0.879	0.877	119	0.869	0.824	150	0.882	0.853	52	0.854	0.904	141	0.851	0.816	59	0.881	0.797	131	0.881	0.840	0.047*
D2S1338	65	0.842	0.831	119	0.844	0.874	150	0.837	0.793	52	0.860	0.904	141	0.837	0.823	59	0.841	0.814	121	0.821	0.339	0.588*
D19S433	65	0.812	0.785	119	0.808	0.832	150	0.813	0.787	52	0.779	0.615	141	0.833	0.738	59	0.822	0.814	131	0.819	0.542	0.339*
VWA	65	0.812	0.846	119	0.808	0.832	150	0.813	0.787	52	0.779	0.615	141	0.833	0.738	59	0.822	0.814	131	0.819	0.542	0.339*
TPOX	65	0.785	0.846	119	0.654	0.672	150	0.649	0.707	52	0.703	0.750	141	0.707	0.681	59	0.868	0.797	132	0.685	0.629	0.082
D18S51	65	0.884	0.892	119	0.874	0.874	150	0.872	0.833	50	0.853	0.820	141	0.877	0.681	59	0.868	0.797	128	0.875	0.805	0.080*
D5S818	65	0.748	0.754	119	0.769	0.773	150	0.751	0.733	52	0.753	0.750	141	0.754	0.759	59	0.759	0.763	133	0.793	0.812	-0.024
FGA	65	0.884	0.892	119	0.866	0.840	150	0.860	0.833	52	0.873	0.904	141	0.862	0.823	59	0.870	0.814	132	0.852	0.811	0.049*
Significant heterozygote deficiency																						

* $P < 0.05$.

10.8 ky,⁴ comparable to the estimated age of M123 STR variation obtained through the method described by Zhivotovsky *et al*⁴⁶ for UAE (11.1 ± 3.9 ky) and Yemen (10.6 ± 4.1 ky), although allelic differences between these two populations indicate that they do not share a common ancestry. Recent archaeological finds supports a trading relationship between Mesopotamia and the Arabian Gulf region dating back to the Al Ubaid Period (~7000 yBP) as evidenced by the excavation of Ubaid pottery from Mesopotamia in UAE.⁸⁻¹⁰ Ancient maritime trade routes linking Mesopotamia to the Indus Valley included Dilmun (the island of Bahrain) and Magan (in the southeastern tip of the Arabian Peninsula). It is possible that the close ties between Mesopotamia with both the Nile River Valley and the ancient Persian Gulf region during the Neolithic helped disseminate these haplogroups.

UAE is characterized by polymorphic levels of E3b1a-M78 (7.9%), similar to the Qatari (4.2%; $\chi^2 = 1.12$, d.f. = 1, $P = 0.29$), with lower values in Oman⁴ (1.7%; $\chi^2 = 5.49$, d.f. = 1, $P = 0.02$) and greater frequencies in Egypt⁴ (18%; $\chi^2 = 6.73$, d.f. = 1, $P = 0.01$) where it is the highest M35 derivative. The majority of the UAE M78 representatives belong to the E3b1a3-V22 clade (6.7%). STR networks of this newly defined marker indicate that it parallels the M78 haplotype cluster δ , although some discrepancies exist.³⁶ Based on the distribution and high STR differentiation of cluster δ , its dispersal may have occurred early, the first to spread the E3b1a-M78 chromosomes to North Africa and then the Near East.⁵⁷

Origin of J1-M267

Previous studies on haplogroup J1-M267 have documented high frequencies of this haplogroup in the areas of Oman (38%),⁴ Iraq (33.1%),²² Egypt (20%),⁴ Lebanon (12.5%)²³ and Turkey (8.99%).¹² The combination of these data with the high frequency of J1-M267 in the Yemeni (72.6%), Qatari (58.3%) and UAE (34.8%) samples examined in the present study reveals a decreasing frequency moving from southern Arabia northwards (Spearman's correlation coefficient with ranks based on distance from Yemen: $r = 0.9286$, $n = 8$, $P < 0.01$). It is also distributed throughout the northwestern African populations at considerable frequencies (35.0 and 30.1% in Algeria and Tunisia, respectively).⁵⁸ Based on binary and STR markers, the greatest degree of differentiation for J1-M267 is detected in the Levant with two distinct demographic dispersals generating its current distribution. A higher observed STR diversity of this clade among Europeans and Ethiopians in comparison to populations of North Africa points to its arrival to Ethiopia and Europe during Neolithic times with a more recent appearance in the latter.⁵⁸ Semino *et al*⁵⁸ describe a YCAIIa22-YCAIIb22 motif in the North African (>90%) and Middle Eastern (>70%) J1-M267 representatives that is less frequent in Ethiopia and Europe, postulating that the dispersal of the M267-YCAIIa22-YCAIIb22

clade occurred during the Arab expansion in the seventh century A.D.

Median BATWING expansion times based on Y-STR data for the Omani (2.3 ky; 95% CI: 0.6–29.2) J1-M267 chromosomes⁴ indicate a more recent arrival to the South Arabian populations as compared to the older expansion times obtained for the Egyptian (6.4 ky; 95% CI: 0.6–278.5)⁴ and Turkish (15.4 ky; 95% CI: 0.4–604.8)¹² representatives of this haplogroup. Conversely, in the present study, Y-STR age estimates based on the method described by Zhivotovsky *et al*⁴⁶ generated much older values for the J1-M267 haplogroup in Yemen, Qatar and UAE (9.7 ± 2.4 , 7.4 ± 2.3 and 6.4 ± 1.4 ky, respectively) than seen in the Omani,⁴ consistent with an earlier arrival to the region during the Neolithic. The data suggest expansion from the north during the Neolithic (or perhaps more recently), which is also reflected in the lower STR variances in southern Arabia (0.14 for Qatar, 0.15 for UAE, 0.20 for Yemen and 0.27 for Oman⁴ versus 0.31 in Egypt⁴ and 0.51 in Turkey¹²). Subsequently, a series of recent demographic events may account for the high haplogroup frequency of J1-M267 in the populations from the present study.

Implications of Y-chromosome distribution in Arabia

Overall, the southern Arabian populations segregate together at an intermediate position with populations from the Levant in the MDS plot, appropriate considering their strategic geographic location at a major bidirectional gateway connecting Africa and Eurasia. Based on the AMOVA, it is also possible to deduce that the overall Y-haplogroup substructure observed in these regions is affected more by geography ($\phi_{ct} = 0.16$) than by language ($\phi_{ct} = 0.08$). Upon classifying the populations based on language family affiliations, the variance among populations within groups is greater ($\phi_{sc} = 0.14$) than the ϕ_{ct} attributable to variation among groups. This difference can be expected since the Afro-Asiatic family encompasses a large variety of languages. Pairwise comparisons of the 32 populations based on Y-haplogroup frequency data (Supplementary Table 1) revealed that only 13 of 90 comparisons display nonsignificant differences within the Afro-Asiatic family, leaving a total of 77 pairwise comparisons generating significant differences. The 13 pairs with nonsignificant differences involve the Levantine populations of Lebanon and Syria possibly as a result of their central position in relation to other Afro-Asiatic groups, whereas the remaining three include the populations within northwest Africa.

Studies focused on this crossroads for human movements have identified geographical barriers that may have limited gene flow with neighboring regions. Specifically, a study based on 15 autosomal STR loci detected a concentration of genetic homogeneity within the Near East, suggesting that the Saharan desert, the Iranian deserts and the Hindu Kush Mountains may have acted as

obstacles for dispersal.¹⁹ The portrayal of the Dasht-e Kavir and Dash-e Lut deserts of Iran as barriers to gene flow has been described in the context of the R1a1-M198 lineage^{14,16–18} as well as in the dissemination of R1b1a-M269 within Iran.¹⁴ Moreover, an admixture analysis by Regueiro *et al*¹⁴ identified the harsh, mountainous terrain in North-east Turkey as well as the Hindu Kush Mountains as limiting factors of gene flow to the Iranian Plateau, whereas the Balochistan acted as a possible conduit for human dispersals. This coastal region that encompasses parts of South Iran, Afghanistan and Pakistan may have provided a unique corridor along the Gulf of Oman.

To examine the degree and geographic extent of genetic homogeneity within the Gulf of Oman, the frequency of the predominant haplogroups were contrasted among the populations in the region. A χ^2 -test on the haplogroup frequencies of Oman, UAE, South Iran¹⁴ and South Pakistan³⁰ indicates that the most frequent haplogroups, E ($\chi^2 = 20.836$, d.f. = 3, $P < 0.0001$), J ($\chi^2 = 8.677$, d.f. = 3, $P = 0.0339$) and R ($\chi^2 = 40.142$, d.f. = 3, $P < 0.0001$) are not evenly distributed among the four populations. As the MDS plot displayed a close affiliation between South Pakistan and North Iran and the former segregated away from the Gulf of Oman populations, the χ^2 -test was repeated excluding South Pakistan. Although statistically significant differences are still apparent for haplogroup E ($\chi^2 = 10.170$, d.f. = 2, $P = 0.0062$) and R ($\chi^2 = 10.560$, d.f. = 2, $P = 0.0051$), J ($\chi^2 = 2.577$, d.f. = 2, $P = 0.2757$) exhibits an even distribution among Oman, UAE and South Iran. However, a greater homogeneity is observed among the South Arabian populations of Oman, UAE and Qatar for haplogroups E ($\chi^2 = 2.249$, d.f. = 2, $P = 0.3248$), J ($\chi^2 = 4.831$, d.f. = 2, $P = 0.0893$) and R ($\chi^2 = 0.308$, d.f. = 2, $P = 0.8573$). The significant differences in frequency of haplogroups result in detectable clines moving from the South Arabian populations to South Iran and then South Pakistan (E: 18.8, 6.8 and 3.3%; J: 50.4, 35.0 and 25.3%; and R: 11.2, 25.6 and 46.2% for South Arabia, South Iran¹⁴ and South Pakistan,³⁰ respectively).

In addition, South Pakistan, South Iran, UAE, Oman and Qatar (although to a lesser extent) share a similar Y-haplogroup substructure with clinal decreases in diversity detected as one moves west to Africa, north to the Levant and Caucasus and east to south and central Asia (Figure 2). Although the Hindu Kush Mountains and Iranian deserts may have played a significant role in encapsulating the region and limiting gene flow,^{14,25} the coastal area may have served as a unique corridor facilitating dispersals into and out of the region at various times in recent human evolution.

At another extreme, the haplogroup distribution of Yemen shows very limited variation, particularly when compared to neighboring populations, Oman and UAE (3 versus 11 haplogroups each), whereas Qatar is intermediate with a total of seven haplogroups, four of which display

frequencies of less than 3.0%. Although Qatar does not approximate the lack of diversity seen in Yemen, the two populations display affinities that are apparent in the MDS plot, in which populations of the Levant are interspersed among the South Arabian populations, with Qatar and Yemen segregating apart from both UAE and Oman.

Regional autosomal STR analysis

To investigate the underlying reasons for the limited Y-chromosome diversity in Yemen, the observed heterozygosity values of 15 highly polymorphic autosomal STR loci were calculated using samples from Kenya,¹⁹ Egypt,¹⁹ Oman,¹⁹ Yemen,¹⁹ Iran,²⁵ Qatar²⁶ and UAE (Cadenas, unpublished results) and are presented in Table 4. Owing to the large number of alleles that exist at each locus, obtaining heterozygote deficiencies may be indicative of a high degree of consanguinity within populations. Qatar possesses 8 out of 15 loci with significant heterozygote deficiency ($P < 0.05$), approximated by Yemen (4 loci) and Oman (3 loci), whereas UAE and Iran display only two loci followed by Egypt and Kenya with one locus each.

A series of recent demographic events may offer an explanation for the Y-haplogroup distribution observed in Yemen. The J1-M267 Y-pattern in particular may have arisen as a result of a founder effect followed by genetic drift. Furthermore, nonrandom-mating practices are common in the area, with cultural beliefs that support polygamy and patrilocal behaviors that perpetuate specific male lines within the region. In addition, consanguineous marriages, particularly among first cousins, are common in the Middle East due to Muslim tradition. This form of inbreeding can serve to propagate a specific patrilineage. Although a combination of these processes probably played a part in forming the Y-haplogroup substructure seen in Yemen, based on the regional autosomal STR analysis, it is likely that inbreeding may have been a significant contributing factor.

A study performed within Sana'a City, Yemen revealed a incidence of consanguinity of 44.7%, with first-cousin marriages comprising 71.6%, and an average coefficient of inbreeding (the probability of an individual having two alleles identical by descent at a given locus) of 0.02442,⁵⁹ almost double that of the Egyptian population (0.01)⁶⁰ and four times that of the Turkish population (0.0064532).⁶¹ Similar studies conducted in Qatar indicate a rate of consanguinity of 54.0% (first cousin marriages accounting for 34.8%) and a coefficient of inbreeding of 0.02706,⁶² whereas comparable consanguinity values were observed in UAE (50.5%)⁶³ and Oman (35.9%).⁶⁴ These figures are representative of the region as a whole where consanguineous marriages are prevalent (28.96% in Egypt,⁶⁰ 33% in Syria,⁶⁵ 51.2–54.4% in Jordan,^{66,67} 57.7% in Saudi Arabia⁶⁸ and 54.4% in Kuwait⁶⁹).

It is significant that in spite of these characteristics, which tend to temper genetic diversity, high Y-chromo-

some haplogroup variability is exhibited in the Gulf of Oman coastal crescent. Patrilineal systems, polygamy and consanguinity are forces that will favor limited diversity along the lines of what is seen in Yemen. It is likely that the region's continued critical role in trade has rendered it an important point of contact between populations and a target of attacks in attempts to gain control of trade from the Persian Gulf. Furthermore, Oman's role in the East African slave trade has been well documented and supported by previous studies^{4,5,54} and may account, at least partially, for the greater diversity it displays.

Conclusion

A comparison of Y-haplogroup substructure of the populations surrounding the Gulf of Oman reveals similarities among them with detectable clines in haplogroup frequencies. This can be attributed to the existence at different times of a coastal corridor along the Gulf of Oman that may have facilitated dispersals into and out of the area. Chromosomes like E3b1c-M123 support archaeological data linking the Fertile Crescent with trading cities along the Persian Gulf, whereas derivatives of E3b1-M35 point to a Neolithic arrival to southern Arabia via the Levant. The limited variability seen in Yemen (and to some extent Qatar) does not mirror the diversity observed in the coastal populations of UAE, Oman, South Iran and South Pakistan. An analysis of heterozygosity using hypervariable autosomal STR loci indicates that both Yemen and Qatar display a deficiency in observed heterozygosity that may be affected to some extent by high rates of consanguineous marriages in the region. In addition, a string of relatively recent events may have maintained Oman and UAE in close contact with other cultures, including attempts to gain control of the Persian Gulf and Oman's involvement in the East African slave trade.

Acknowledgements

We gratefully acknowledge Tenzin Gayden and Maria Regueiro for their contributions to this study as well as Sheyla Mirabal for her technical assistance.

References

- 1 Lahr MM, Foley R: Multiple dispersals and modern human origins. *Evol Anthropol* 1994; 3: 48–60.
- 2 Stringer C: Coasting out of Africa. *Nature* 2000; 405: 24–25,27.
- 3 Oppenheimer S: *Out of Eden: The Peopling of the World*. Constable: London, 2003.
- 4 Luis JR, Rowold DJ, Regueiro M et al: The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 2004; 74: 32–544.
- 5 Rowold DJ, Luis JR, Terreros MC, Herrera RJ: Mitochondrial DNA gene flow indicates preferred usage of the Levant corridor over the Horn of Africa. *J Hum Genet* 2007; 52: 436–447.
- 6 Hassan F: Archeology and linguistic diversity in North Africa; In: Bellwood P, Renfrew C (eds): *Examining the Farming/Language*

- Dispersal Hypothesis*. McDonald Institute for Archaeological Research: Cambridge, 2002, pp 127–133.
- 7 Militarev A: The prehistory of a dispersal: the proto-Afrasian (Afroasiatic) farming lexicon; In Bellwood P, Renfrew C (eds): *Examining the farming/language dispersal hypothesis*. McDonald Institute for Archaeological Research: Cambridge, 2002, pp 135–150.
 - 8 Flavin K, Shepherd E: Fishing in the Gulf: preliminary investigations at an Ubaid site, Dalma (UAE). *PSAS* 1994; **24**: 115–134.
 - 9 Beech M, Elders J: An 'Ubaid-related settlement on Dalma Island, Abu Dhabi Emirate, United Arab Emirates. *Bull Soc Arabian Stud* 1999; **4**: 17–21.
 - 10 Beech M, Cuttler R, Moscrop D *et al*: New evidence for the Neolithic settlement of Marawah Island, Abu Dhabi, United Arab Emirates. *PSAS* 2005; **35**: 37–56.
 - 11 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton University Press: Princeton, 1994.
 - 12 Cinnioglu C, King R, Kivisild T *et al*: Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 2004; **114**: 127–148.
 - 13 Nasidze I, Ling EYS, Quinque D *et al*: Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann Hum Genet* 2004; **68**: 205–221.
 - 14 Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ: Iran: Tricontinental nexus for Y-driven migration. *Hum Hered* 2006; **61**: 132–143.
 - 15 Quintana-Murci L, Chaix R, Wells RS *et al*: Where west meets east: the complex mtDNA landscape of the southwest and central Asian corridor. *Am J Hum Genet* 2004; **74**: 827–845.
 - 16 Quintana-Murci L, Krausz C, Zerjal T *et al*: Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 2001; **68**: 537–542.
 - 17 Wells RS, Yuldasheva N, Ruzibakiev R *et al*: The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 2001; **98**: 10244–10249.
 - 18 Qamar R, Ayub Q, Mohyuddin A *et al*: Y chromosomal DNA variation in Pakistan. *Am J Hum Genet* 2002; **70**: 1107–1124.
 - 19 Shepard EM, Herrera RJ: Genetic encapsulation among Near Eastern populations. *J Hum Genet* 2006; **51**: 467–476.
 - 20 Kivisild T, Reidla M, Metspalu E *et al*: Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet* 2004; **75**: 752–770.
 - 21 Sanchez JJ, Hallenberg C, Borsting C, Hernandez A, Morling N: High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. *Eur J Hum Genet* 2005; **13**: 856–866.
 - 22 Al-Zahery N, Semino O, Benuzzi G *et al*: Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* 2003; **28**: 458–472.
 - 23 Semino O, Passarino G, Oefner PJ *et al*: The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 2000; **290**: 1155–1159.
 - 24 Thomas MG, Parfitt T, Weiss DA *et al*: Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba-the 'Black Jews of Southern Africa'. *Am J Hum Genet* 2000; **66**: 674–686.
 - 25 Shepard EM, Herrera RJ: Iranian STR variation at the fringes of biogeographical demarcation. *For Sci Int* 2005; **158**: 140–148.
 - 26 Perez-Miranda AM, Alfonso-Sanchez MA, Pena JA, Herrera RJ: Qatari DNA variation at a crossroad of human migrations. *Hum Hered* 2006; **61**: 67–79.
 - 27 Arredi B, Poloni ES, Paracchini S *et al*: A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 2004; **75**: 338–345.
 - 28 Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA: Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 2002; **70**: 265–268.
 - 29 Nasidze I, Sarkisian T, Kerimov A, Stoneking M: Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Hum Genet* 2003; **112**: 255–261.
 - 30 Sengupta S, Zhivotovskiy LA, King R *et al*: Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* 2006; **78**: 202–221.
 - 31 Kivisild T, Rootsi S, Metspalu M *et al*: The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 2003; **72**: 313–332.
 - 32 Karafet T, Xu L, Du R *et al*: Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 2001; **69**: 615–628.
 - 33 Antunez de Mayolo G, Antunez de Mayolo A, Antunez de Mayolo P *et al*: Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis* 2002; **23**: 3346–3356.
 - 34 Underhill PA, Passarino G, Lin AA *et al*: The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001; **65**: 43–62.
 - 35 Shen P, Lavi T, Kivisild T *et al*: Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum Mutat* 2004; **24**: 248–260.
 - 36 Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R: Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a STR-network-based approach through six new biallelic markers. *Hum Mutat* 2006; **27**: 831–832.
 - 37 Martinez L, Reategui EP, Fonseca LR *et al*: Superimposing polymorphism: the case of a point mutation within a polymorphic *Alu* insertion. *Hum Hered* 2005; **59**: 109–117.
 - 38 Hammer MF, Horai S: Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 1995; **56**: 951–962.
 - 39 Y Chromosome Consortium: A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; **12**: 339–348.
 - 40 Reynolds J, Weir BS, Cockerham CC: Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* 1983; **105**: 767–779.
 - 41 SPSS for Windows, Rel. 11.0.1.2001. SPSS Inc: Chicago.
 - 42 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.
 - 43 Schneider S, Roessli D, Excoffier L: A software for population genetics data analysis. Arlequin version 2.000. Genetics and Biometry Laboratory, University of Geneva: Switzerland, 2000.
 - 44 Carmody G: *G-test*. Carleton University: Ottawa, 1991.
 - 45 Kimura M, Ohta T: Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Natl Acad Sci USA* 1978; **75**: 2868–2872.
 - 46 Zhivotovskiy LA, Underhill PA, Cinnioglu C *et al*: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004; **74**: 50–61.
 - 47 Cruciani F, Santolamazza P, Shen P *et al*: A back migration from Asia to sub-saharan Africa is supported by high resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 2002; **70**: 1197–1214.
 - 48 Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB: Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 1994; **91**: 3166–3170.
 - 49 Kittles RA, Perola M, Peltonen L *et al*: Dual origins of Finns revealed Y chromosome haplotype variation. *Am J Hum Genet* 1998; **62**: 1171–1179.
 - 50 Kayser M, Krawczak M, Excoffier L *et al*: An extensive analysis of Y-chromosomal STR haplotypes in globally dispersed human populations. *Am J Hum Genet* 2001; **68**: 990–1018.
 - 51 Zhivotovskiy LA, Goldstein DB, Feldman MW: Genetic sampling error of distance ($\delta\mu$)² and variation in mutation rate among STR loci. *Mol Biol Evol* 2001; **18**: 2141–2145.

- 52 Raymond M, Rousset F: GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 1995; **86**: 248–249.
- 53 Rousset F, Raymond M: Testing heterozygote excess and deficiency. *Genetics* 1995; **140**: 1413–1419.
- 54 Richards M, Rengo C, Cruciani F *et al*: Extensive female-mediated gene flow from sub-Saharan Africa into Near Eastern Arab populations. *Am J Hum Genet* 2003; **72**: 1058–1064.
- 55 Olivieri A, Achilli A, Pala M *et al*: The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* 2006; **314**: 1767–1770.
- 56 Underhill PA, Shen P, Lin AA *et al*: Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000; **26**: 358–361.
- 57 Cruciani F, La Fratta R, Santolamazza P *et al*: Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 2004; **74**: 1014–1022.
- 58 Semino O, Magri C, Benuzzi G *et al*: 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* 2004; **74**: 1023–1034.
- 59 Gunaid AA, Hummad NA, Tamim KA: Consanguineous marriage in the capital city Sana'a, Yemen. *J Biosoc Sci* 2004; **36**: 111–121.
- 60 Hafez M, El-Tahan H, Awadalla M, El-Khayat H, Abdel-Gafar A, Ghoneim M: Consanguineous mating in the Egyptian population. *J Med Genet* 1983; **20**: 58–60.
- 61 Basaran N, Sayli BS, Basaran A, Solak M, Artan S, Stevenson JD: Consanguineous marriages in the Turkish population. *Clin Genet* 1988; **34**: 339–341.
- 62 Bener A, Alali KA: Consanguineous marriage in a newly developed country: The Qatari population. *J Biosoc Sci* 2005; **38**: 239–246.
- 63 Al-Gazali LI, Bener A, Abdulrazzaq YM, Micallef R, Al-Khayat AI, Gaber T: Consanguineous marriages in the United Arab Emirates. *J Biosoc Sci* 1997; **29**: 491–497.
- 64 Rajab A, Patton MA: A study of consanguinity in the Sultanate of Oman. *Ann Hum Bio* 2000; **27**: 321–326.
- 65 Prothro ET, Diab LN: *Changing Family Patterns in the Arab East Beirut*. American University: Beirut, 1974.
- 66 Khoury SA, Massad D: Consanguineous marriage in Jordan. *Am J Med Genet* 1992; **43**: 769–775.
- 67 Sueyoshi S, Ohtsuka R: Effects of polygyny and consanguinity on high fertility in the rural Arab population in South Jordan. *J Biosoc Sci* 2003; **35**: 513–526.
- 68 El-Hazmi MAF, Al-Swailem AR, Warsy AS, Al-Swailem AM, Sulaimani R, Al-Meshari AA: Consanguinity among Saudi Arabian population. *J Med Genet* 1995; **32**: 623–626.
- 69 Al-Awadi SA, Moussa MA, Naguib KK *et al*: Consanguinity among the Kuwait population. *Clin Genet* 1985; **27**: 483–486.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)