

Iran: Tricontinental Nexus for Y-Chromosome Driven Migration

M. Regueiro^a A.M. Cadenas^a T. Gayden^a P.A. Underhill^b R.J. Herrera^a^aDepartment of Biological Sciences, Florida International University, Miami, Fla., and ^bDepartment of Genetics, Stanford University, Stanford, Calif., USA

Key Words

Iran · Y-chromosome · SNP

Abstract

Due to its pivotal geographic position, present day Iran likely served as a gateway of reciprocal human movements. However, the extent to which the deserts within the Iranian plateau and the mountain ranges surrounding Persia inhibited gene flow via this corridor remains uncertain. In order to assess the magnitude of this region's role as a nexus for Africa, Asia and Europe in human migrations, high-resolution Y-chromosome analyses were performed on 150 Iranian males. Haplogroup data were subsequently compared to regional populations characterized at similar phylogenetic levels. The Iranians display considerable haplogroup diversity consistent with patterns observed in populations of the Middle East overall, reinforcing the notion of Persia as a venue for human disseminations. Admixture analyses of geographically targeted, regional populations along the latitudinal corridor spanning from Anatolia to the Indus Valley demonstrated contributions to Persia from both the east and west. However, significant differences were uncovered upon stratification of the gene donors, including higher proportions

from central east and southeast Turkey as compared to Pakistan. In addition to the modulating effects of geographic obstacles, culturally mediated amalgamations consistent with the diverse spectrum of a variety of historical empires may account for the distribution of haplogroups and lineages observed. Our study of high-resolution Y-chromosome genotyping allowed for an in-depth analysis unattained in previous studies of the area, revealing important migratory and demographic events that shaped the contemporary genetic landscape.

Copyright © 2006 S. Karger AG, Basel

Introduction

Assessment of the current state of genetic distributions in regions of the Middle East provides an indication of past settlements and movements in the area. The genetic diversity that exists throughout this vast region [1–4] points to its central role in the dispersal and expansion of human populations throughout the tri-continental area. The genetic diversity of the people that populate the area is often affected by the regional geography, and while certain geographic features, such as the deserts of Iran, have served as barriers, while others, such as the Strait of Bab el Mandab, have acted as conduits for population movements [5, 6]. Furthermore, environmental fluctuations occurring over time have changed areas that once served

M. Regueiro and A.M. Cadenas contributed equally to the article.

as a passageway to a barrier as in the case of the Strait of Hormuz that connects the Arabian Peninsula to present day Iran. Migration routes linking Africa and Eurasia via the Levantine corridor [1, 4, 7] and the Horn of Africa [1, 5–8] have been described previously. Dispersals associated with the Horn of Africa are facilitated via the Strait of Babel Mandab (Gate of Sorrow) that connects the Red Sea with the Indian Ocean through the Gulf of Aden. Yet, the routes and crossings beyond the Gate of Sorrow involving the Arabian Peninsula are typically neglected in genetic analyses. The Strait of Hormuz is a narrow, shallow waterway sprinkled with islands bisecting the Persian Gulf and the Gulf of Oman. Sea level changes that have exposed the Gulf floor during particular periods in time would have provided a natural bidirectional gateway for populations expanding to the east from the Arabian Peninsula to the Iranian plateau as well as in the opposite direction.

Two major deserts, the Dasht-e Kavir and Dash-e Lut in Iran, have also had an influence in the peopling of Southwest Asia. The Dasht-e Kavir occupies the north central region of present day Iran while to the east the Dash-e Lut extends north to south. The breadth of these two deserts encompasses a majority of the central Iranian plateau and its uninhabitable conditions may have led to the preferential settlement of populations to the surrounding areas. Their portrayal as geographical barriers to gene flow has previously been supported by the Y-chromosome distribution of the R-M198 lineage [9–11] as well as mitochondrial DNA analysis [12]. However, high-resolution genetic analysis within Iran has been lacking and we address this deficiency here.

Unlike most other systems, the genetic diversity within the Y-chromosome shows a high level of correlation with geography making it a particularly informative set of loci to evaluate gene flow, especially when data sets are taken to an appropriate level of phylogenetic molecular resolution. Regional high-resolution studies of Y-chromosome polymorphisms within the Iranian population should provide insights on human dispersals due its key geographic location at a nexus of Africa, Asia and Europe. Previous investigations of Y-chromosome variation in Southwest Asia have focused on several prominent regions. For example, the Y-haplogroup distribution in the Caucasus displayed a closer relationship to West Asia than Europe and also suggested no evidence that the Caucasus Mountains acted as pronounced barriers to gene flow [3] while Y-chromosome analysis in Anatolia revealed considerable haplotype variation identifying Turkey as both source and recipient of gene flow [2]. On the other hand, Luis et al. [4] employed Y-specific single-nucleotide and

short tandem repeat polymorphisms to assess the relative prominence of the Levant versus the Horn of Africa in the bidirectional dispersals involving Africa and Eurasia and revealed a propensity towards the spread of signature Neolithic markers through the Levantine corridor with limited gene flow via the Horn of Africa. Two previous studies of Y-chromosome haplogroup structure in Iran that did not attain the level of resolution of the present study have been carried out. Specifically, one study apportioned the gene pool to 10 haplogroups (E, F, J2, K, L, O3, P, R1, R1a1 and R2) deep within the Y-chromosome genealogy [10] while the second associated the Dravidian and Indo-European language dispersals with the demic diffusion of both early farmers from southwestern Iran and pastoral nomads from west and central Asia into India based on the frequencies of haplogroups J and R1a1 in 401 samples [9]. Despite this preliminary work, an evaluation of the extent of Anatolian and Indus Valley influence within the central Persian contact zone of this broad corridor would require a greater haplogroup resolution. In this study, high-resolution Y-chromosome analyses of 150 males from Iran were utilized to generate a comprehensive haplogrouping of contemporary diversity to elucidate the role Iran played in population dispersal across the latitudinal belt spanning western Anatolia to the Indus Valley. The Iranian haplogroup profile was also considered in relation to data from Oman to evaluate the role of the Strait of Hormuz as a potential barrier to gene flow. Additionally, information obtained was analyzed in conjunction with data previously published from other regional, geographically targeted regions. This data was then interpreted within the context of geographical barriers in order to envision the extent to which these features either acted as deterrents or facilitated gene flow.

Materials and Methods

Collection and Extraction of DNA Samples

Blood samples from 150 Iranian males were collected in EDTA Vacutainer tubes. Donors provided information on their paternal ancestry for at least two previous generations. DNA was extracted from the blood using the phenol-chloroform extraction method as previously described [13]. Information on the geography, ethnic origin, and linguistic affiliation of the populations is provided in table 1. Ethical guidelines were adhered to as stipulated by the institutions involved.

Y-Haplogroup Analysis

A total of 97 binary genetic markers [2, 7, 14, 15] were genotyped by standard methods, including PCR/RFLP, allele-specific PCR [16] and the YAP polymorphic *Alu* insertion (PAI) [17]. The

Table 1. Description of the populations included in the study

Geographic region	Population	Abbreviations	n ^a	Reference	Language	
					family	sublevel
North Africa	Arabs (Algeria)	AA	35	37	Afro-Asiatic	Semitic
	Berbers (Algeria)	BA	19	37	Afro-Asiatic	Berber
	Tunisia	TN	148	37	Afro-Asiatic	Semitic
	Egypt	EG	147	4	Afro-Asiatic	Semitic
East Africa	Kenya	KE	29	4	Niger-Congo	Benue-Congo, Bantu
	Somalia	SO	201	38	Afro-Asiatic	Cushitic
	Oromo	OR	78	39	Afro-Asiatic	Cushitic
	Amhara	AM	48	39	Afro-Asiatic	Semitic
Arabian peninsula	Oman	OM	121	4	Afro-Asiatic	Semitic
Caucasus	Georgia	GE	77	24	South Caucasian	Georgian
	Armenia	AR	100	24	Indo-European	Armenian
	Azerbaijan	AZ	72	24	Altaic	Turkic
Near East or Levant	Lebanon	LE	31	27	Afro-Asiatic	Semitic
	Syria	SY	20	27	Afro-Asiatic	Semitic
	Iraq	IQ	139	29	Afro-Asiatic	Semitic
South Asia	Pakistan		176	15	Indo-European	Indo-Iranian
	North	NP	85			
	South	SP	91			
	India			28		
	Punjab	PU	66		Indo-European	Indo-Iranian
	Gujarat	GU	29		Indo-European	Indo-Iranian
	Konkanastha	KB	43		Indo-European	Indo-Iranian
	Brahmin					
Chenchu	CH	41		Dravidian	South Central	
Koya	KO	41		Dravidian	South Central	
Iranian plateau	Central Iran/Esfahan	ES	50	24	Indo-European	Indo-Iranian
	Iranians in					
	Uzbekistan	UZ	53	10	Indo-European	Indo-Iranian
	North Iran	NI	33	present study	Indo-European	Indo-Iranian
South Iran	SI	117	present study	Indo-European	Indo-Iranian	
South Europe/Anatolia	Greece	GR	76	27	Indo-European	Greek
	Turkey		523	2	Altaic	Turkic
	North	NT	112			
	Middle	MT	172			
	South	ST	76			

^a Number of individuals sampled.

DNA fragments were separated by 1X TAE, 3% agarose gel electrophoresis and scored subsequent to ethidium bromide staining and UV light photography. The Y-SNP haplogroup nomenclature used follows the recommendations of the Y Chromosome Consortium [18].

Statistical and Phylogenetic Analyses

The Iranian samples were segregated into a northern (n = 33) and a southern (n = 117) collection, with the line of demarcation

running along the geographic position of the Dash-e Lut and Dasht-e Kavir deserts. The Dasht-e Kavir desert bisects northern Iran from southern Iran, while the Dash-e Lut desert in eastern Iran further segregates these two regions. In addition to the Iranian samples, twenty-five reference populations from throughout the Middle East region reported in previous studies (table 1) were included in the statistical and phylogenetic analyses to obtain a thorough representation of Y-haplogroup variation and phylogeographic relationships. Two sets of tests of molecular variance

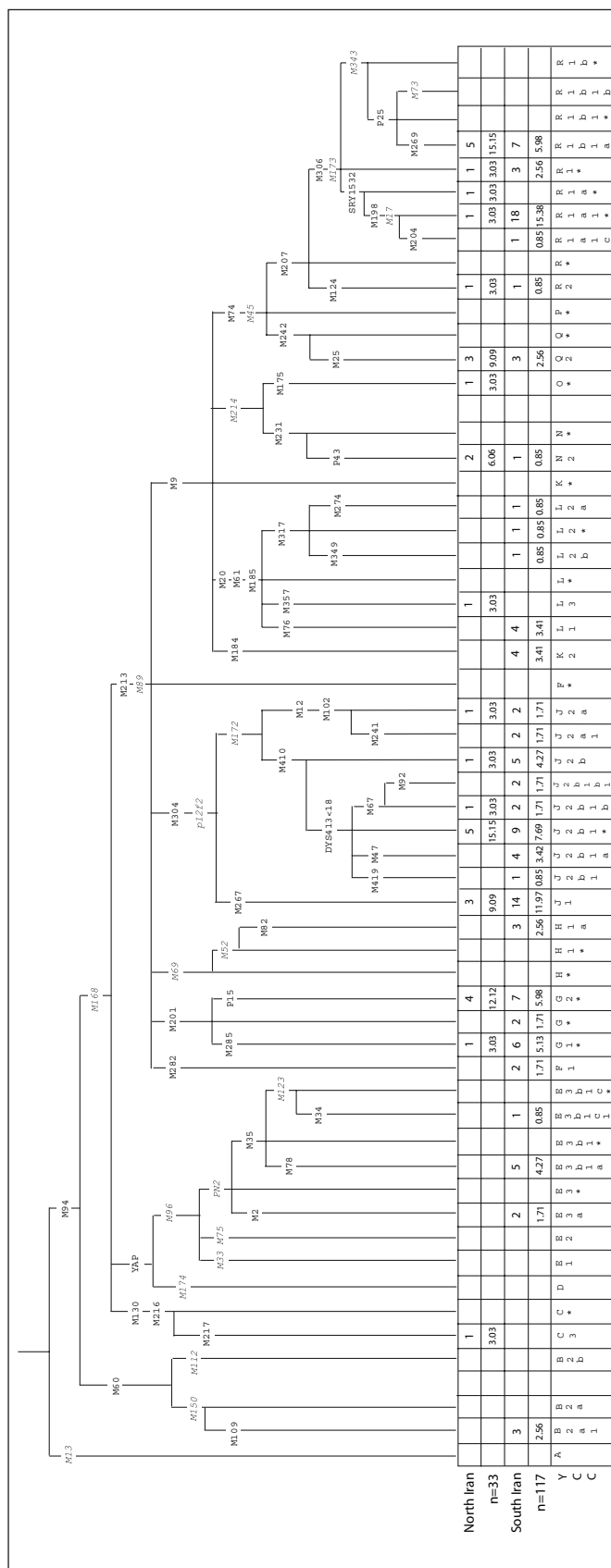
(AMOVA) [19] were conducted with the Arlequin version 2.000 package using the 27 populations with subdivisions of groups based on geography (North Africa, East Africa, Arabian Peninsula, Caucasus, Near East/Levant, South Asia, Iranian Plateau and Southeast Europe/Anatolia), linguistic family (Afro-Asiatic, Indo-European, Niger-Congo, Dravidian, Altaic and South Caucasian) and linguistic sublevel (Semitic, Berber, Benue-Congo, Georgian, Armenian, Turkic, Indo-Iranian, South Central and Greek). Maximum-likelihood (ML) phylogeny was constructed using the Phylip v3.6 program [20]. A correspondence analysis was performed using the NTSYSpc-2.02i package by Rohlf [21]. Admixture proportions (mY) were calculated with the program Admix 2.0 [22] utilizing Turkey [2] and Pakistan [15] as the parental populations and Iran (present study) as the hybrid population. The parental populations were systematically subdivided by region in order to determine the full extent of their contribution to the hybrid population. The population from Pakistan was subdivided into a north and a south region. The subdivisions for Turkey were made in reference to the demarcations assigned in figure 1 of Cinnioglu et al. [2], in which the Turkish population was divided into nine regions and statistically significant latitudinal and longitudinal differences were detected. Based on the existence of these differences and for the purpose of this analysis particular regions were grouped together. Regions 3, 4 and 5 comprised east Turkey while west Turkey consisted of the remaining six regions. In addition, a northeast (regions 2 and 3), central east (regions 4 and 7) and southeast (regions 5 and 6) subdivision were also made. The contribution for each set of parental populations examined was analyzed using all of Iran as well as the subdivided regions of northern and southern Iran.

Results

Phylogeography

Analysis of 97 biallelic markers from the Y chromosome of 33 northern Iranians and 117 southern Iranians identified 34 paternal lineages that are distributed throughout haplogroups B, C, E, F, G, H, J, K, L, N, O, Q, and R. The phylogenetic relationships and frequencies of the 34 Y-chromosome lineages observed in Iran are illustrated in figure 1 according to YCC nomenclature [18].

Fig. 1. Phylogenetic relationships of Y-chromosome haplogroups and frequencies observed for the populations analyzed. The 19 markers shown in italics were not genotyped and are included to provide context. Lineages with more than one marker listed are homologous. The following 44 markers were genotyped but not observed: M37, M39, M56, M65, M68, M77, M93, M97, M99, M119, M122, M128, M136, M137, M153, M154, M156, M157, M158, M163, M167, M191, M197, M222, M258, M268, M280, M286, M289, M290, M318, M319, M320, M321, M327, M339, M340, M365, M368, M369, and M377.



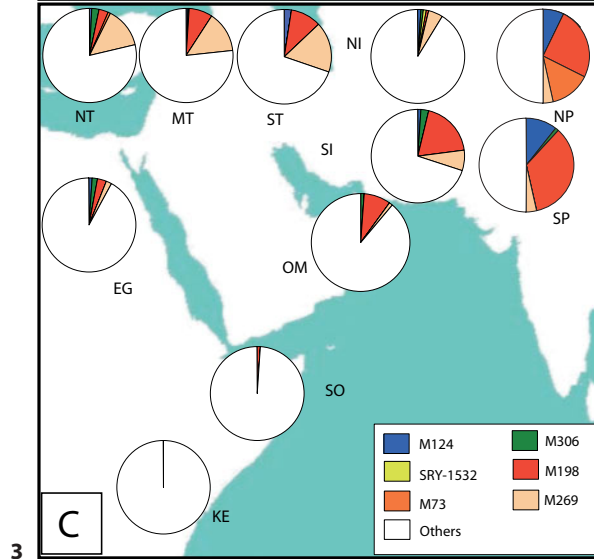
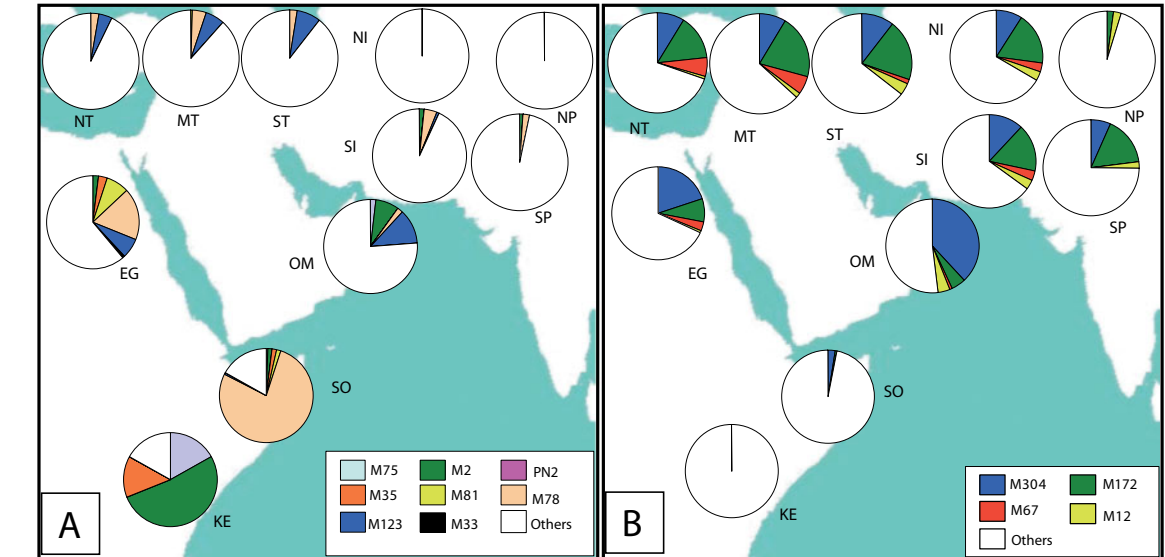
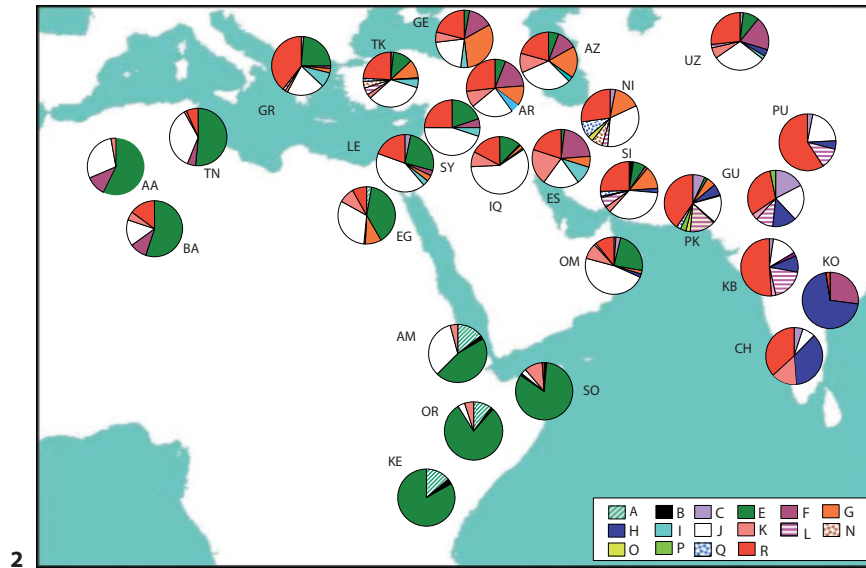


Fig. 2. Geographic distribution of major Y-chromosome haplogroup frequencies for the 27 populations listed in table 1.
Fig. 3. Geographic distribution of Y-chromosome haplogroup frequencies in selected populations. Population abbreviations can be found in table 1. (A) E clade, (B) J clade, and (C) R clade.

Table 2. Results of AMOVA

	Variance, %	Φ_{ST}	Φ_{SC}	Φ_{CT}
No grouping:				
Among populations	17.04			
Within populations	82.96	0.17*		
Geography				
Among groups	13.22			0.13*
Among populations within groups	4.67		0.05*	
Within population	82.11	0.18*		
Language (family)				
Among groups	9.41			0.09**
Among populations within groups	9.51		0.10*	
Within population	81.08	0.19*		
Language (sublevel)				
Among groups	14.47			0.14*
Among populations within groups	4.14		0.05*	
Within population	81.39	0.18*		
Φ_{ST} = Variance within populations; Φ_{SC} = variance among populations within groups; Φ_{CT} = variance among groups. * $p < 0.001$; ** $p < 0.002$.				

In northern Iran, group J represents the most frequent haplogroup (33.33%) followed by groups R (27.27%), G (15.15%), Q (9.09%) and N (6.06%). Similarly, in southern Iran group J exhibits the highest frequency (35.04%) and is trailed by groups R (25.64%), G (12.82%), E (6.84%), L (5.98%) and K (3.42%). Figure 2 displays the major haplogroup distributions for the 27 populations listed in table 1, while the geographic distribution of informative clades E, J and R and their derivatives in Iran and nine reference populations typed at a similar resolution are illustrated in figure 3.

It is interesting to note that haplogroup E representatives are found solely in the southern region of Iran. More specifically, E3a-M2 is present in only two individuals (1.71%) while the frequency for clade E3b1-M35 is higher (5.12%). The subclade E3b1a-M78 accounts for the majority of the Iranian representatives of E3b1-M35.

The J2b1b*-M67 lineage, a subclade of M172, exhibits a frequency of 3.03% in northern Iran and 1.71% in the south. Its sister subclade, J2b1b1-M92, is present only in southern Iran (1.71%). Other informative distributions within haplogroup J include J2a-M12 present in north and south Iran (3.03 and 3.42%, respectively).

Haplogroup K contains the remaining downstream groups of L through R in the two Iranian groups. Haplogroup R comprises most of the northern and southern Iranian samples that carry the K-M9 mutation (27.27 and

25.64%, respectively). The majority of these group R samples fall within R1-M306 while a single individual from each region of Iran belongs to the R2-M124 lineage. In north Iran, individuals within the R1-M306 clade can be further subdivided into R1-M306*, R1a1*-M198, R1b1a-M269 and R1a*-SRY1532 (XM198) occurring with frequencies of 3.03, 3.03, 15.15 and 3.03%, respectively. Individuals from south Iran that comprise the R1-M306 clade segregate into R1-M306* (2.56%), R1a1*-M198 (15.38%), R1b1a-M269 (5.98%), and R1a1c-M204 (0.85%).

AMOVA

Table 2 displays the results of the AMOVA analysis based on the northern and southern Iranians examined in this study in addition to the 25 reference populations. The highest fraction of the variability as expected resides within populations but there was a substantial percentage due to differences among populations (17.04%, $p < 0.001$) indicating a high level of population differentiation. Assignment of the populations according to eight geographical groups generated an Φ_{ct} (variance among groups) of 0.13 ($p < 0.001$) and an Φ_{sc} (variance among populations) of 0.05 ($p < 0.001$), indicating significant inter- and intraregional geographical structuring. Variation among groups and among populations within groups correlates significantly with linguistic partitioning when the populations

were grouped according to language family with an Φ_{ct} of 0.09 ($p < 0.002$) and an Φ_{sc} of 0.10 ($p < 0.001$). Further division of the populations according to linguistic sublevels did not generate significantly different values for both among groups [Φ_{ct} of 0.14 ($p < 0.001$)] and among populations within groups [Φ_{sc} of 0.05 ($p < 0.001$)].

Phylogenetic Analyses

Segregation of populations along biogeographical lines and their genetic affinities is illustrated in the ML analysis (fig. 4). For this analysis, all data sets were normalized to common levels of molecular resolution. There is an apparent geographic structuring observed with distinct regions of the Middle East and adjacent areas represented within five main clusters (South Asia, Caucasus, Sub-Saharan Africa, North Africa and Levant/Arabian Peninsula). The dendrogram places the populations of north and south Iran intermediate between south Asia on one side and the Levant, Caucasus, Arabian Peninsula and Africa on the other. The populations within India display an expected position in relation to nearby Iran, with Gujarat (north-west India) exhibiting a closer affiliation to the groups from Pakistan and the Iranian Plateau. Punjabis and Konka Brahmins on one branch, and the Koyas and Chenchus on the other segregate terminally within the clade. In addition, there is an apparent contrast in the segregation of peoples from the Iranian plateau. North and south Iran from the present study cluster with Turkey while Esfahan (Central Iran) and the Iranians exiled in Uzbekistan display a higher affinity to the inhabitants of the Caucasus. It is notable that the correspondence analysis of the 27 populations and 16 haplogroups (fig. 5) corroborate the ML phylogeny.

Admixture Analysis

Table 3 displays admixture proportions (mY) calculated for Iran, north Iran, and south Iran. These proportions are based on various parental population combinations that resulted from the subdivision of Pakistan and Turkey by regions. In the case of north Iran, a contribution of 100% from the east and 0% from the west is observed in all instances. In contrast, south Iran displays a more heterogeneous contribution upon subdivision of the parental populations. South Iran exhibits a greater proportion of south Pakistani (0.659) than Turkish (0.341) genetic influence but a smaller proportion of north Pakistani (0.417) than Turkish (0.583) influence. In addition, when the data from Turkey is partitioned, the proportion of genes contributing to the genetic pool of south Iran is higher for east Turkey (0.429) than west Turkey (0.267), with the Paki-

Table 3. Admixture proportions (mY) for Iran, North Iran, and South Iran based on various parental population combinations

Parental populations	Hybrid population		
	Iran	North Iran	South Iran
Pakistan (mY1)	0.40	0.00	0.67
Turkey (mY2)	0.60	100.00	0.33
N. Pakistan (mY1)	0.25	0.00	0.42
Turkey (mY2)	0.75	100.00	0.58
S. Pakistan (mY1)	0.40	0.00	0.66
Turkey (mY2)	0.61	100.00	0.34
Pakistan (mY1)	0.23	0.00	0.57
E. Turkey (mY2)	0.77	100.00	0.43
Pakistan (mY1)	0.51	0.00	0.73
W. Turkey (mY2)	0.49	100.00	0.27
Pakistan (mY1)	0.52	0.00	0.74
N.E. Turkey (mY2)	0.48	100.00	0.27
Pakistan (mY1)	0.07	0.00	0.45
C.E. Turkey (mY2)	0.93	100.00	0.55
Pakistan (mY1)	0.14	0.00	0.48
S.E. Turkey (mY2)	0.86	100.00	0.52

North Pakistan, South Pakistan, Northeast Turkey, Central east Turkey and Southeast Turkey were compared in all possible permutations but did not yield significant results. mY1 = Admixture proportion for parental population 1; mY2 = admixture proportion for parental population 2.

stani component greater in each case (0.571 and 0.733, respectively). Of note is the decrease in the proportion of northeast Turkey's contribution to south Iran (0.265) in comparison to southeast and central east Turkey (0.520 and 0.552, respectively). North Pakistan, south Pakistan, northeast Turkey, central east Turkey and southeast Turkey were compared in all possible permutations but did not yield significant results.

Discussion

In the present study, high-resolution Y-chromosome analyses revealed a relatively high diversity of haplogroups in Iran, with individuals from the southern region distributed throughout 11 groups and those in the northern region within 8 groups. In the ML phylogeny Iran occupies an intermediary position, with north and south Iran proximal to populations of Anatolia and the Caucasus, suggesting that it served as a bridge between South Asia, Africa and Arabia (fig. 4). This observation is corroborated further by the correspondence analysis, where popula-

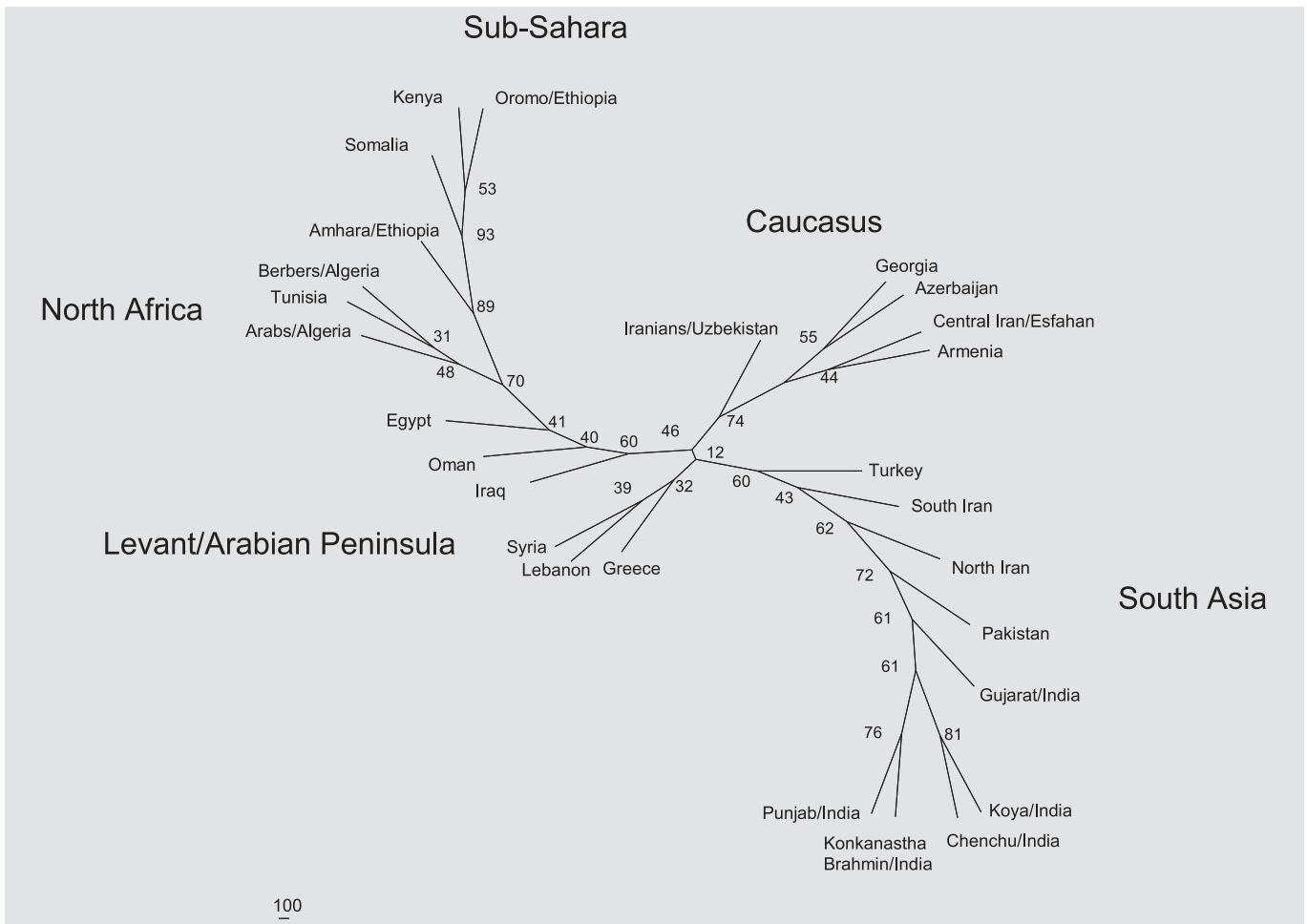


Fig. 4. ML radial phylogeny based on the Y-SNP haplogroup frequency data of 27 populations listed in table 1.

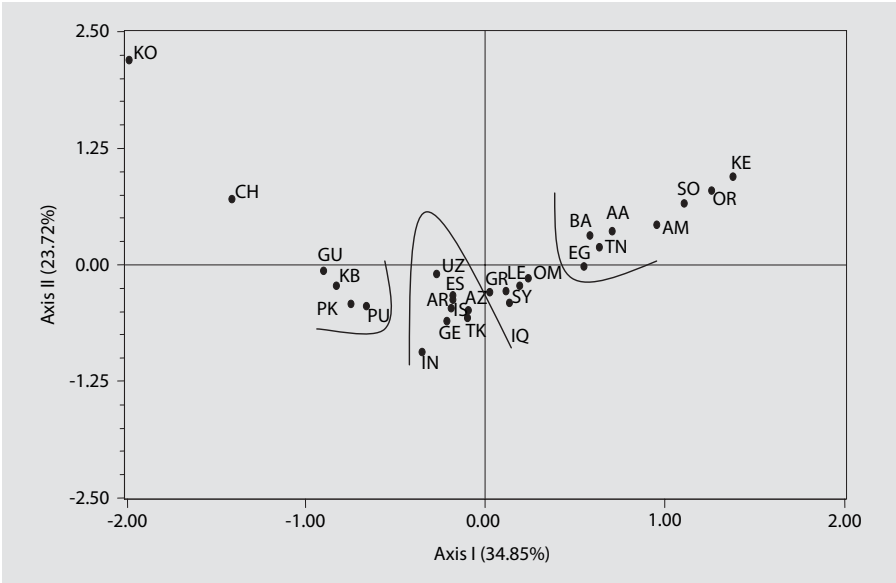


Fig. 5. Correspondence analysis of Y-haplogroup frequency data from 27 populations. Population abbreviations can be found in table 1. The curves represent guidelines to indicate the major divisions between groups of populations.

tions encompassing the Iranian plateau provide a connection between these regions (fig. 5). Studies performed with other marker systems (mtDNA and autosomal STRs) examining populations in this region reiterate the trend of high diversity observed by the patrilineal Y chromosome [12, 23].

Iran's Role as a Crossroad for Bidirectional Gene Flow and the Impact of the Iranian Deserts as Inhibitors of Y-Chromosome Dispersal

Compared to one another, the populations of north and south Iran display significant differences in haplogroup distributions as well as in relation to the Esfahan of Central Iran [24] and the Iranians of Uzbekistan [10]. The correspondence analysis grouped these four populations within the same cluster (fig. 5); however, the ML phylogeny segregated these four populations into two clades (fig. 4). North and south Iran remain together with populations of Anatolia, while Esfahan and the Iranians of Uzbekistan grouped with populations from the Caucasus. Although admixture with Uzbekistan natives may explain why the exiles from Iran cluster with populations from the Caucasus, it is not apparent why the Esfahan segregate away from the north and south Iranians of this study.

The NRY haplogroup distribution within Iran displays considerable diversity, with the partitioning of particular lineages (i.e. M198, M269) between the northern and southern regions consistent with the existence of geographic barrier(s) that limited gene flow among these areas. The Dash-e Lut desert in eastern Iran and the Dasht-e Kavir desert in north-central Iran represent areas of harsh conditions where habitation and even migration would have been difficult. The partitioning of Y-haplogroups within Iran is particularly evident when the distribution of specific lineages of Asian, African and European origins are examined.

A group R lineage of Asian descent, R1a1-M198, is found at higher frequencies in the south rather than the north of Iran (15.38 vs. 3.03%). Interestingly, the M198 lineage has been linked with the spread of the Kurgan culture originating in southern Russia/Ukraine in the region between the Black and Caspian Seas [9, 10, 25]. Furthermore, Wells et al. [10] described a reduction in frequency of M198 eastward across Siberia to the Altai-Sayan Mountains and Mongolia, and southward into India. This finding supports the inferred migration of the Indo-Iranians during the period 3,000 to 1,000 B.C. as proposed by Mallory [26], although models of one haplogroup coinciding with one migration are inherently simplistic, especially

when high frequency clades within contiguous continental areas are involved. In fact it has been shown that at least some R1a1 chromosomes dispersed earlier [15]. Additional molecular resolution within R1a1 will be required to dissect potential independent overlying subsequent events associated with this widespread ancient haplogroup. From the disparate M198 frequencies observed for the north and south of Iran, it is possible to envision a movement southward towards India where the lineage may have had an influence on the populations south of the Iranian deserts and where the Dash-e Lut desert would have played a significant role in preventing the expansion of this marker to the north of Iran. The lower frequencies of M198 in the region of Anatolia (11.8% in Greece [27] and 6.9% in Turkey, with a statistically significant longitudinal correlation [2]) and the Caucasus (10% in Georgia, 6% in Armenia and 7% in Azerbaijan) [24] suggests that population movement was southward towards India and then westward across the Iranian plateau. In addition, the detection of rare R1-M173* and R1a-SRY1532 lineages in Iran at higher frequencies than observed for either Turkey, Pakistan or India suggests the hypothesis that geographic origin of haplogroup R may be nearer Persia.

Both major lineages associated with group E, M2 and M35, were detected in the southern portion of Iran (6.8%). E3a-M2 is hypothesized to have dispersed south from northern Africa within the last 3,000 years by the Bantu agricultural expansion [30–32] while E3b1-M35 is thought to have originated in East Africa [33]. Unfortunately the number of individuals segregated into each lineage is so low that it does not allow questions about routes exiting Africa to be addressed, yet it is clear that populations from this region had an impact on the Iranian gene pool.

The Iranian corridor may have been vital in the dissemination of groups from the west. Haplogroup J1-M267 characterizes African and Arabian populations, with a decreasing frequency northwards. Previous studies have documented relative high frequencies for this lineage in the areas of Oman (38%) [4], Iraq (33%) [29], Egypt (20%) [4], Lebanon (13%) [27], and Turkey (9%) [2]. The lineage is represented in Iran at similar levels for the north (9.09%) and south (11.97%). J2-M172 is believed to have expanded from the Fertile Crescent to Europe during the demic diffusion of Neolithic farmers [27]. J2 and its derivatives are equally represented in the neighboring populations of Iran (23%) and Turkey (22.2%) [2]. Its influence in northern Iran (24.2%) is comparable to southern Iran (23%). A decrease in the frequency of J2 can be seen east of the Iranian Plateau in south Pakistan (18.7%), with a dramatic decline in north Pakistan (4.7%) [15].

Another R lineage, R1b1a-M269, is found at high frequencies in populations from Europe [34]. There is a decrease in frequency moving to the southeast, with 14.5% in Turkey [2], approximately 11% in Iraq [29], 2 and 1% in Egypt and Oman respectively [4], 8% in Iran, and 2.8% in Pakistan [15]. In contrast to J2, this marker is found in the northern region at a higher frequency than in the south (15.15 vs. 5.98%, respectively), suggesting preferential expansion to northern Iran as the population moved from east to west.

The admixture analysis also conveyed a parallel message regarding the extent to which the deserts of Iran serve as barriers to gene flow. The population of north Iran consistently displays a maximum proportion of Turkish markers (100%) and a minimum proportion of Pakistani genes (0%) in their gene pool (table 3). This significant observation exists regardless of Pakistan's geographically closer proximity to Iran, indicating a substantial geographic barrier preventing gene flow to the region from the east.

Anatolian and Indus Valley Influences in Persia

Population admixture proportions from Turkey and Pakistan revealed considerable information regarding how much these populations contributed to the haplogroup substructure within Iran. The admixture proportions obtained from north Iran were discussed in the previous section. In contrast, data obtained using south Iran revealed an influence from both the east and west, with Pakistan comprising a larger proportion of the south Iranian gene pool compared to Turkey (0.669 vs. 0.331). However, upon subdividing the parental populations by regions, interesting variations in the admixture proportions were observed. The contribution from south Pakistan is greater than that of north Pakistan, which is to be expected when taking into consideration the geography of the region. The northern region of Pakistan is bordered along the west by the Hindu Kush Mountains, serving as a formidable wall to gene flow westward. On the other hand, an arid, coastal region on the Iranian plateau named the Balochistan that encompasses parts of present day Iran, Pakistan, and Afghanistan would provide a natural route for gene flow between south Pakistan and south Iran. Furthermore, a significant genetic heterogeneity is observed within Pakistan and these differences can be generalized to populations of the north and south, each showing greater contributions from Asia and Africa, respectively [35].

Analysis of the admixture proportions from east and west Turkey also generated expected results with a greater

contribution from the eastern region most likely due to its closer proximity to Iran and the heavy southeastern European penetrations mainly from present day Greece. However, dividing Anatolia into northeast, central east, and southeast regions revealed a significant under representation in admixture proportions for the northeast (0.265, 0.552, and 0.520, respectively). The geography of Turkey is predominantly mountainous, including the Taurus Mountains, extending diagonally along the eastern border of Anatolia and a series of ranges in the north along the Black Sea. In the center of the peninsula stretching diagonally from the central west towards the southeast is the Anatolian plateau. Population expansions from the central Anatolian plateau towards the southern coast in the direction of the Fertile Crescent and Iran could explain the larger influence observed from these two regions, with populations of northeast Turkey encountering more obstacles in the form of harsh, mountainous terrain and thus contributing less to the gene pool of the southern Iranians. It is possible that although the distance from Anatolia to northern Iran is shorter than the one provided by the southern corridor to Persia, the bountiful land of the Tigris-Euphrates may have been favored. Moreover, populations in the northern region of Anatolia are subject to greater gene flow from Eastern Europe via the maritime routes facilitated by the Black Sea such as the southeastern European captives imported into northern Anatolia to form the Ottoman troops known as the Janissaries [36].

Regardless of any regional subdivisions, the admixture analysis indicates that both Pakistan and Turkey have contributed to the gene pool of Iran (0.402 and 0.598, respectively). In spite of Pakistan's closer proximity, geographical barriers may have been formidable enough to hinder gene flow. Furthermore, other factors could have aided gene exchange as well and contributed to Turkey's larger contribution to Iran despite the greater physical distance. Throughout history, major dominions have encompassed and thus united the regions of Anatolia and Persia, including the Babylonian (1800–1170 BC), Assyrian (1170–612 BC), Persian (550–323 BC), and Ottoman (1350–1918 AD) Empires. These dynasties in essence acted as cultural facilitators of gene flow.

Conclusion

The Iranians from the present study carry signature lineages with origins in Asia, Africa, and Europe. There is an observable difference between north and south Persia with the southern region possessing a slightly greater va-

riety of NRY lineages. This finding can be explained to some extent by the surrounding geographical features (the Dasht-e Kavir to the south, the Dash-e Lut to the southeast, the Hindu Kush mountain ranges to the east), resulting in gene flow limitations. Although surrounded with these obstacles, north Iran continues to exhibit a greater spectrum of haplogroups than its neighboring populations.

As a whole, Persia displays contributions from both Anatolia and the Indus Valley, yet a greater proportion from the former is observed in spite of Iran's proximity to the latter. A combination of cultural and geographical fa-

cilitators may be responsible for the stronger genetic influence of Turkey on Iran. For example, culturally a continuous string of empires has amalgamated the territories currently occupied by Turkey and Iran effectively bringing about some degree of genetic homogeneity. Geographically, the fertile corridor along the Tigris and Euphrates rivers may have aided in population dispersals from Anatolia. Overall, Iran's strategic location created a point of contact for many human migrations. And Persia's role as a crossroad for tricontinental dispersals is today reflected in its diversity of haplogroups and lineages.

References

- Cavalli-Sforza LL, Menozzi P, Piazza A: The history and geography of human genes. Princeton University Press, Princeton, NJ, 1994.
- Cinnioğlu C, King R, Kvisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA: Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 2004;114:127–148.
- Nasidze I, Ling EYS, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O, Sarraf-Zadegan N, Naderi GA, Asgari S, Sardas S, Farhud DD, Sarkisian T, Asadov C, Kerimov A, Stoneking M: Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann Hum Genet* 2004;68:205–221.
- Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ: The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 2004;74:32–544.
- Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavy K, Santachiara-Benerecetti AS: Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* 1999;23:437–441.
- Stringer C: Coasting out of Africa. *Nature* 2000;405:24–25, 27.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL: The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001;65:43–62.
- Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R: Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 1999;9:1331–1334.
- Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K: Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 2001;68:537–542.
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Bing S, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF: The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 2001;98:10244–10249.
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ: Y chromosomal DNA variation in Pakistan. *Am J Hum Genet* 2002;70:1107–1124.
- Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Mehdi SQ, Torroni A, McElreavey K: Where west meets east: the complex mtDNA landscape of the southwest and central Asian corridor. *Am J Hum Genet* 2004;74:827–845.
- Antunez de Mayolo G, Antunez de Mayolo A, Antunez de Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ: Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis* 2002;23:3346–3356.
- Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman MW, Oefner PJ: 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.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CT, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA: 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.
- Martinez L, Reategui EP, Fonseca LR, Sierra-Montes JM, Terreros MC, Pereira-Simon S, Herrera RJ: Superimposing polymorphism: The case of a point mutation within a polymorphic *Alu* insertion. *Hum Hered* 2005;59:109–117.
- Hammer MF, Horai S: Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 1995;56:951–962.
- Y Chromosome Consortium: A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002;12:339–348.
- 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.
- Felsenstein J: PHYLIP-phylogeny inference package (version 3.2). *Cladistics* 1989;5:164–166.
- Rohlf F: NTSTSp. Exter Publishing, Setauket, NY, 2002.
- Dupanloup I, Bertorelle G: Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol* 2001;18:672–675.
- Shepard EM, Herrera RJ: Iranian STR variation at the fringes of biogeographical demarcation. *For Sci Int* 2005;158:140–148.
- 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.

- 25 Passarino G, Semino O, Magri C, Al-Zahery N, Benuzzi G, Quintana-Murci L, Andell-novic S, Bullc-Jakus F, Liu A, Arslan A, Santachiara-Benerecetti AS: The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. *Hum Immunol* 2001;62:922–932.
- 26 Mallory JP: In search of the Indo-Europeans: Language, archaeology and myth. Thames and Hudson, London, 1989.
- 27 Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill P: The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. *Science* 2000; 290:1155–1159.
- 28 Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R: The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 2003;72:313–332.
- 29 Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torrioni A, Santachiara-Benerecetti AS: 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.
- 30 Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL: Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 1998;15:427–441.
- 31 Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS: Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 1998;62:420–434.
- 32 Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torrioni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A: Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 1999;65:829–846.
- 33 Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torrioni A, Cavalli-Sforza LL, Underhill P, Santachiara-Benerecetti AS: 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.
- 34 Cruciani F, Santolamazza P, Shen P, Macauley V, Moral P, Olckers A, Modiano D, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torrioni A, Cavalli-Sforza LL, Scozzari R, Underhill P: 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.
- 35 Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ: Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 2002; 70:1107–1124.
- 36 'Janissaries.' The Columbia Encyclopedia, 6th ed. New York: Columbia University Press, 2001–04. Retrieved November 14, 2005, from www.bartleby.com/65.
- 37 Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C: A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 2004;75:338–345.
- 38 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.
- 39 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.