

Erica M. Shepard · Rene J. Herrera

Genetic encapsulation among Near Eastern populations

Received: 11 December 2005 / Accepted: 26 January 2006 / Published online: 24 March 2006
© The Japan Society of Human Genetics and Springer-Verlag 2006

Abstract This report aims to genetically characterize the relationships between geographically targeted human populations covering an expanse from east sub-Saharan Africa northeastward into northern India with an emphasis on the Near East. A number of parameters of population genetics interest were examined based on allele frequencies from 15 forensic autosomal STR markers [D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA]. The phylogenetic analyses generated from genetic profiles of 885 individuals indicate that populations west of and including Iran have experienced substantial gene flow. Accordingly, our findings delineate a region of genetic homogeneity concentrated within the Near East with increasing genetic differentiation moving south into Africa and further east into Asia. We suggest that the Saharan desert, the Hindu Kush mountain range and perhaps to a lesser extent, the deserts of Iran may have acted as southern, eastern and northern geographical barriers, respectively, forming a genetic enclosure that allows limited gene flow outside the Near East. The biparental genetic landscape supports a picture of close contact between the Arab and Persian populations, perhaps beginning during the initial settlement of Asia from Africa extending to recent times.

Keywords Autosomal · STRs · Allele · Phylogeny · Population genetics

E.M. Shepard · R.J. Herrera (✉)
Department of Biological Sciences,
Florida International University, University Park,
OE 304, Miami, FL 33199, USA
E-mail: herrerar@fiu.edu
Tel.: +1-305-3481258
Fax: +1-305-3481259

Present address: E.M. Shepard
Counterterrorism Forensic Science Research Unit,
FBI Academy, Bldg. 12, Quantico, VA 22135, USA

Introduction

The origins of anatomically modern humans can be traced back to Africa as early as 200,000 years before present (YBP) evidenced by many archeological, linguistic and genetic data, the whole of which renders the “Out of Africa” hypothesis a generally accepted scientific precept (Watson et al. 1997; Quintana-Murci et al. 1999; Underhill et al. 2000; Ke et al. 2001; Maca-Meyer et al. 2001). The genetic composition of Africa has been described as extremely diverse (Chen et al. 1995; Ingman et al. 2000; Jorde et al. 2000; Watkins et al. 2001; Salas et al. 2002), in part due to this early rise of mankind, but also due to complex demic expansions both out of and back into the continent (Hammer et al. 1998; Cruciani et al. 2002; Semino et al. 2002; Maca-Meyer et al. 2003). These waves of migrations and undoubtedly back-migrations occurring at various times throughout human history bestow a unique importance upon East and Northeast Africa. Specifically, the Levantine region in the Nile River delta and the Horn of Africa near the Strait of Sorrow have been proposed as major corridors into Arabia and vice versa (Cavalli-Sforza et al. 1994; Kivisilid et al. 1999; Quintana-Murci et al. 1999; Stringer 2000; Underhill et al. 2000; Bar-Yosef 2002; Nebel et al. 2004; Kivisilid et al. 2004; Luis et al. 2004). Furthermore, Arabia represents the strategic multi-directional thruway intersecting three continents (Europe, Asia and Africa) and four major linguistic families (Caucasian, Elamite, Indo-European and Afro-Asiatic) (Renfrew 1996, 2000). The northern ridge of Arabia, also known as the Fertile Crescent, was the birthplace of agriculture some 8,000 YBP. Additionally, the initial domestication of animals associated with the practice of nomadic pastoralism followed sometime later to the east (Zvelebil 1980). Thus, this region as a focal point of early technological advances of civilization provided the impetus for major demic diffusions permeating in all directions (Quintana-Murci et al. 2001). In terms of extant populations, this report aims to investigate

geographically targeted groups covering an expanse from east sub-Saharan Africa northeastward into northern India with an emphasis on the Near East.

Although Neolithic agriculture arose independently in up to as many as nine different areas (Diamond and Bellwood 2003), the spread of farming and animal domesticates to Europe (approx. 8,000 YBP), North Africa (approx. 7,000 YBP) and further east into Asia (approx. 6,000 YBP) was generally attributed to have Near Eastern origins based on linguistic and archeological evidence (Hassan 2002; Militarev 2002). Considering the recent timescale of these pervasive landmark events in human prehistory, hypervariable genetic markers may offer the finely tuned resolution needed to characterize demic movements stemming from this cradle of civilization. Autosomal short tandem repeat (STR) loci are an informative markers for determining short-term reconstructive phylogenies (Bowcock et al. 1994; Jorde et al. 1995; Bosch et al. 2000; Lum et al. 2002) especially among closely related groups of individuals (Rowold and Herrera 2003; Shepard and Herrera 2005). Accordingly, this study employed 15 tetranucleotide STRs to generate allelic frequencies from

a total of 885 individuals from eleven contemporary, anthropologically well-defined and geographically targeted populations from sub-Saharan Africa (Kenya, Rwandan Hutus, and Sudan), Southwest Asia and the Levant (Bahrain, Egypt, Jordan, Georgia, Oman, and Yemen) and South Asia (Pakistan and Indian Punjabis). Additionally, eight relevant datasets from previously published works from Cabinda (Angola), Mozambique, Tutsis (Rwanda), Iran, Japan, Taiwan, Belgium, and US (Caucasian dataset) were utilized as reference populations to provide information from another 2,173 individuals that were included in several inter-population assessments (Table 1).

Upon examination of these 15 highly polymorphic loci, we find that analysis of the distribution of genetic variation of these 19 populations reveal a strong overall correlation to both language and geography among groups of populations. Also, phylogenetic examinations show genetic segregation/affinities that roughly follow geographic and linguistic clines. Populations from Arabia exhibit a greater degree of genetic homogeneity compared to groups in adjacent regions. Overall, the genetic landscape paints a picture of

Table 1 Description of populations examined

Population	<i>n</i>	Description	Language Affiliation ¹	Geographical Coordinates* ²	Reference
Africa					
Cabinda (Angola)	110	General population	Niger-Congo/Benue-Congo/Bantu	–	Beleza et al. (2004)
Egypt	119	Arab/Berber/Egyptian population	Afro-Asiatic/Semitic/Central	31°00'N; 30°00'E	Present study
Hutu (Rwanda)	95	Hutu population, Central Rwanda	Niger-Congo/Benue-Congo/Bantu	2°00'S; 30°00'E	Present study
Kenya	65	Bantu population	Niger-Congo/Benue-Congo/Bantu	2°00'S; 37°30'E	Present study
Mozambique	142	General population from Maputo	Niger-Congo/Benue-Congo/Bantu	–	Alves et al. (2004)
Sudan	65	General population	Afro-Asiatic/Semitic/Central	16°00'N; 30°00'E	Present study
Tutsi (Rwanda)	124	Tutsi population, Central Rwanda	Niger-Congo/Benue-Congo/Bantu	–	Reguiero (2004)
Asia					
Bahrain	50	Arab population	Afro-Asiatic/Semitic/Central	26°50'N; 50°50'E	Present study
Georgia	65	General population	Kartvelian/Georgian	42°00'N; 43°30'E	Present study
Iran	150	General population	Indo-European/Indo-Iranian/Iranian	33°00'N; 55°00'E	Shepard and Herrera (2005)
Japan	526	General population	Japanese	–	Hashiyada et al. (2003)
Jordan	46	Arab population	Afro-Asiatic/Semitic/Central	32°00'N; 37°00'E	Present study
Oman	141	Arab population	Afro-Asiatic/Semitic/Central	21°00'N; 57°00'E	Present study
Pakistan	94	General population	Indo-European/Indo-Iranian/Indo-Aryan	30°00'N; 70°00'E	Present study
Punjabi (India)	86	Punjab in western India	Indo-European/Indo-Iranian/Indo-Aryan	32°00'N; 76°30'E	Present study
Taiwan	597	General population	Sinto-Tibetan/Chinese	–	Wang et al. (2003)
Yemen	59	Arab population	Afro-Asiatic/Semitic/Central	16°00'N; 48°00'E	Present study
Europe					
Belgium	222	General population, majority from Flanders region	Indo-European/Germanic/West	–	Decorte et al. (2004)
North America					
Caucasian (USA)	302	General population from US	Indo-European/Germanic/West	–	Butler et al. (2003)

*Missing data denoted by (–) indicates information not reported in reference datasets

¹URL for linguistic affiliation web resource: <http://www.ethnologue.com>

²URL for geographical coordinate web resource: <http://www.nationalgeographic.com/xpeditions/atlas/index.html>

geographic barriers to gene flow encapsulating a region centered in Arabia extending from Northeast Africa to Persia. This may suggest a basin of close contact among the human groups within this area. Increasingly differentiated populations are encountered beyond the barriers, south of the Sahara into Africa and further east of Iran into the remainder of Asia. Although geographical obstacles to gene exchange are not a novel concept in population genetics, the collection of human groups in this study provides an opportunity to empirically examine these phenomena. The implications of these findings are discussed within the context of other genetic evidence.

Materials and methods

Population information and sample collection

The eleven populations examined in the present study include the following: sub-Saharan Africa (Kenya, Rwandan Hutus, and Sudan), Southwest Asia and the Levant (Bahrain, Egypt, Jordan, Georgia, Oman, and Yemen) and South Asia (Pakistan and Indian Punjabis) (Table 1). Their geographic locations are indicated in Fig. 1. Data from eight additional worldwide populations were obtained from the literature and employed for comparison purposes in two analyses (Table 1 includes reference database citations). Individuals from each population were identified by biogeographical information gathered and traced back at least two generations. Informed consent was obtained from each individual prior to collection. Each collection was arranged through the community leaders and/or elders of each region and supervised by the same. Sample collections were performed according to the ethical guidelines indicated by Florida International University's Institutional Review Board (IRB).

DNA isolation, PCR amplification and detection of STRs

All samples from these 11 populations were collected as whole blood in EDTA Vacutainer™ tubes. Genomic DNA was extracted by standard phenol-chloroform and ethanol precipitation method (Sambrook and Russell 2001). The samples were amplified by PCR using the commercial AmpFISTR Identifier kit (Applied Biosystems, Foster City, CA, USA) at the following loci: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, FGA, and Amelogenin. Amplifications were carried out in a GeneAmp PCR System 9600 thermocycler (Applied Biosystems, Foster City, CA, USA) with the following cycling parameters: 11 min denaturation at 95°C; 28 cycles of 1 min denaturation at 94°C, 1 min primer annealing at 59°C and 1 min primer extension at 72°C; and a final soak for 60 min at 60°C. A portion of each amplified sample was mixed in formamide and GS500 LIZ as an internal size standard as recommended by the manufacturer (Applied Biosystems, Foster City, CA, USA), and then amplicons were separated using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) CE instrument. GeneScan® 3.7 was employed to determine the fragment sizes and Genotyper® 3.7 NT software was utilized to designate alleles by comparison with the allelic ladder provided by the manufacturer.

Statistical and phylogenetic analysis

Allelic frequencies of the 15 STR loci for 11 populations were calculated by the gene counting method (Li 1976). The Arlequin software package Version 2.000 (Levene 1949; Guo and Thompson 1992; Schneider et al. 2000) was used to assess Hardy–Weinberg equilibrium (HWE)

Fig. 1 Biogeographical area studied



using Fisher's exact test with the modified Markov-chain Monte Carlo method, as well as to determine Nei's gene diversity index (GD) (Nei 1987). The following parameters of population genetics interest were examined: matching probability (MP), power of discrimination (PD), polymorphic information content (PIC) and power of exclusion (PE) using PowerStats program Version 1.2 (Jones 1972; Brenner and Morris 1990; Tereba 1999).

In two analyses Neighbor-Joining (NJ) phylogeny and Multi-Dimensional Scaling (MDS), in addition to the eleven populations datasets from Iranians (Shepard and Herrera 2005) and Rwandan Tutsis (Regueiro 2004) were included due to their regional geographic relevance. Allelic frequencies of the groups studied were employed to generate the NJ tree based on *Fst* distances using the PHYLIP 3.52c software (Felsenstein 2002). Bootstrap consensus scores (1,000 replications) were generated by the SEQBOOT and GENDIST options of the PHYLIP software, while the CONSENSE programs determined the best-fit tree. MDS analysis was performed using the Statistical Package for the Social Sciences (SPSS) software program to summarize multivariate genetic relationships among the 13 groups.

Genetic structuring was analyzed among all 19 populations according to both geographic proximity and linguistic subfamily affiliation through hierarchical analysis of molecular variance (AMOVA) (Excoffier et al. 1992) to examine potential partitioning along these lines on a global scale.

Results

Intra-population STR diversity

Allelic frequency distributions for eleven populations from Africa (Egypt, Hutus from Rwanda, Kenya, and

Sudan) and Asia (Bahrain, Georgia, Jordan, Oman, Pakistan, Punjabis from northwest India, and Yemen) are presented for the first time as part of this study and are available as electronic supplementary materials (Tables 1 through 11 at the following URL: http://www.fiu.edu/~herrerar/Frequency_data.htm). In addition, important population genetics parameters are summarized in Table 2 for each group under study. The combined matching probability (CMP) for these eleven populations ranges from 1 in 6.922×10^{15} in the Jordanian dataset to 1 in 2.032×10^{17} in the Omani dataset. The combined power of exclusion (CPE) ranges from 0.99998973 in the Pakistanis to 0.99999985 in the Hutu from Rwanda. Each of the eleven datasets generate a combined power of discrimination (CPD) value > 0.9999999999999999 . Table 2 lists the loci in each population that do not meet Hardy-Weinberg equilibrium (HWE) expectations when $P < 0.05$ (14 loci out of 165 possible tests). However, after the application of the Bonferroni adjustment ($\alpha = 0.05/15$ or 0.0033) only three loci persist in their departure from HWE: D16S539 and D5S818 in the Pakistani dataset and vWA in the Punjabi collection.

Inter-population STR diversity

To ascertain the genetic relationships between the 13 African and Asiatic populations, a NJ tree was generated using *Fst* distances based upon the allelic frequencies of 15 STR loci. Figure 2 illustrates the NJ tree based on the 11 populations presented in this study for the first time plus Iran and the Tutsis from Rwanda. Within the overall topology, 4 out of the 11 bootstrap values are below 50% incidence. There are roughly three major clusters of populations within the dendrogram (clockwise from the bottom): (1) African/Bahrain/Oman, (2) Jordan/Egypt/Yemen and (3) Georgia/Iran with the two South Asian populations from Pakistan and Punjab.

Table 2 Parameters of population genetics interest

Population	Total alleles	CMP	CPE	CPD	Avg. <i>He</i>	Departures from HWE
Africa						
Egypt	140	1 in 1.018×10^{17}	0.99999855	†	0.7922	None
Hutu (Rwanda)	154	1 in 9.458×10^{16}	0.99999985	†	0.7993	None
Kenya	139	1 in 4.095×10^{16}	0.99999980	†	0.8025	TH01; D5S818
Sudan	132	1 in 7.289×10^{16}	0.99999911	†	0.8014	D18S51
Asia						
Bahrain	127	1 in 5.182×10^{16}	0.99999946	†	0.8051	None
Georgia	125	1 in 9.143×10^{15}	0.99999900	†	0.7855	None
Jordan	121	1 in 6.922×10^{15}	0.99999854	†	0.7977	D2S1338; D19S433
Oman	142	1 in 2.032×10^{17}	0.99999632	†	0.7990	TH01
Pakistan	136	1 in 7.317×10^{16}	0.99998973	†	0.7944	TH01; D13S317; D16S539*; D5S818*
Punjabi (India)	134	1 in 3.104×10^{16}	0.99999873	†	0.7933	D19S433; vWA*
Yemen	125	1 in 1.399×10^{16}	0.99999474	†	0.7907	D13S317; D18S51

CMP combined matching probability, CPE combined probability of exclusion, CPD combined probability of discrimination, Avg. *He* average heterozygosity, HWE hardy-weinberg equilibrium

†Indicates each of the eleven populations generated a Combined Probability of Discrimination > 0.9999999999999999

*Indicates persistence of HWE departure after Bonferroni-like adjustment for number of loci tested ($\alpha: 0.05/15 = 0.0033$)

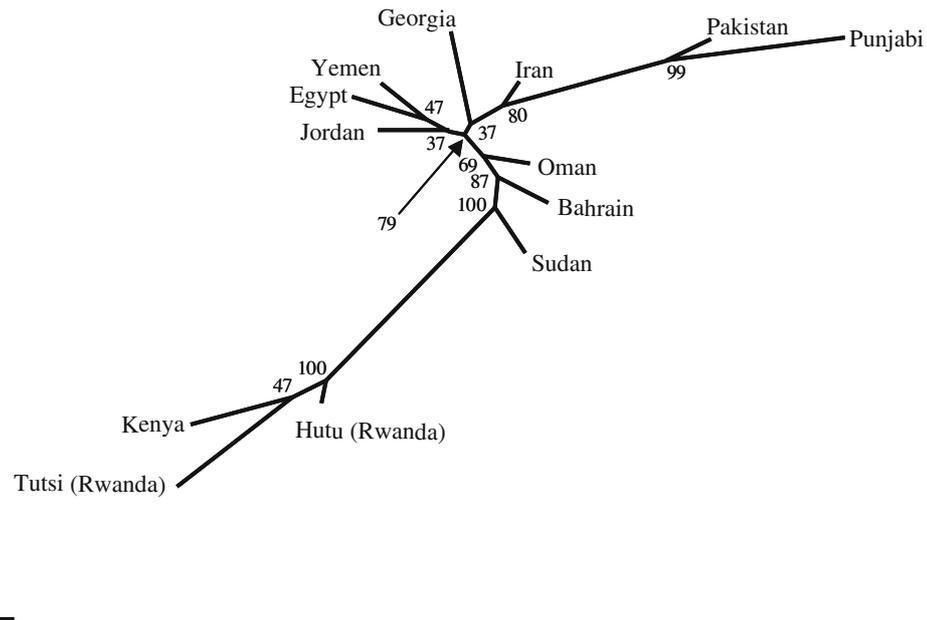


Fig. 2 Neighbor-Joining (*NJ*) phylogenetic analyses of 13 populations based on STR allelic frequencies using *Fst* distances. The GENDIST option of the Phylip software created branch distances

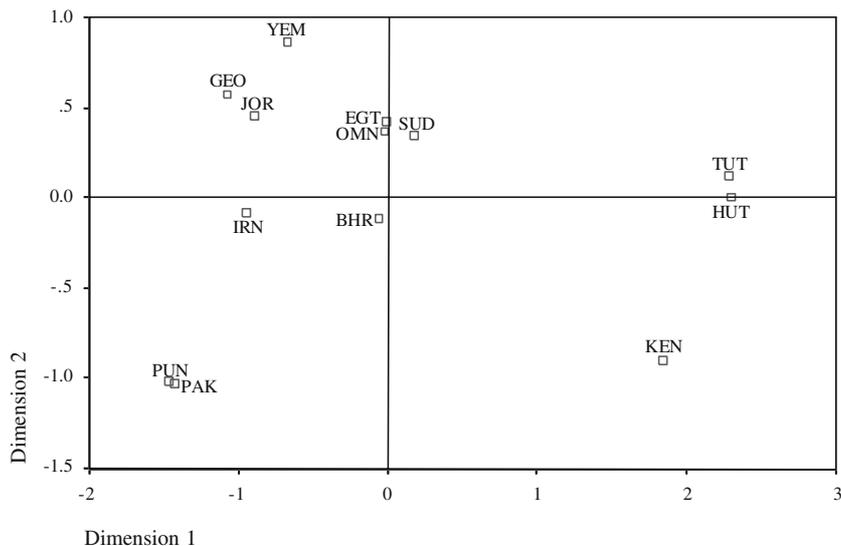
onto which the corresponding bootstrap values (based on 1,000 replications) were transferred to the corresponding nodes of the NJ tree

The three sub-Saharan groups (Hutus, Tutsis and Kenyans) cap the end of the African/Arab cluster distantly from the nearest bifurcation from Sudan. Similarly, in the second clade, the groups of Pakistan and Punjab segregate far off from Iran and Georgia. It is interesting to note that 8 of the 13 populations (Sudan, Bahrain, Oman, Iran, Georgia, Jordan, Egypt, and Yemen) fall in close proximity to the trifurcation (bootstrap value 79%) of these three clusters. Oman segregates singly from the main African/Arab cluster initially (69%), followed next by Bahrain (87%). The Sudanese are distant from the main clustering of continental African groups, bifurcating much closer to the former two Arab populations (100%). Within that African cluster, the two Rwandan populations, the Hutu and Tutsi, do not segregate together. Instead, the Hutus bifurcate first (100%) followed by the Tutsi and Kenyan groups at the extreme of the branch. The length of the Tutsi branch is long relative to the other immediate populations indicating distinct genetic differentiation. Moving in a counter-clockwise direction to the next cluster, Georgia initially separates from the remaining populations (37%) followed by Iran (80%), then more distantly by Pakistan from the Punjabis from India. Lastly, Jordan initially separates from Egypt and Yemen (37%), while the latter two groups segregate from each other with a 47% bootstrap value.

A MDS analysis based on *Fst* distances was performed using these same 13 populations to examine the phylogenetic relationships based on the 15 STR loci (Fig. 3). The layout of the two-dimensional map in terms of spacing of the populations is consistent with the overall topology of the NJ tree. The proportion of

variance accounted for by the corresponding distances of the scaled data is 95.67%. The main differences in this particular analysis from the phylogram are with regard to the proximities of the populations to each other within their respective main clusters. For example, while the sub-Saharan African populations fall in a linear arrangement together to the right side within the upper and lower right quadrants, the Hutu and Tutsi populations from Rwanda are closer in proximity to each other while the Kenyans remain distant. Similarly to the Hutu and Tutsi, the Pakistan and Punjabi groups are in closer proximity than in the NJ analysis. The same eight groups that agglutinated together surrounding the trifurcation of the dendrogram fall into or in close proximity to each other in the upper left quadrant of the MDS plot. However, the arrangement is slightly different, for instance, while the small distance between Sudan and Oman is repeated, Bahrain is farther away and Egypt is now closer to the former two groups. Both Georgia and Jordan fall in the middle of the upper left quadrant with Yemen close to the upper left limit of the graph. Both Iran and Bahrain lie in singular positions within the lower left quadrant just below the boundary that divides the upper from the lower quadrants with Bahrain closer to the crosshairs of the plot. The same two phylogenetic analyses (NJ and MDS) were performed with all 19 groups including the eight reference populations. The topology of both the NJ tree and MDS plot mirror the ones based on the 13 populations described above. One notable difference is that Georgia segregated in the same clade as Belgium and US Caucasians instead of within the Asian cluster (data is not shown).

Fig. 3 Multi-dimensional scaling (MDS) analyses of 13 populations using *Fst* distances based of STR allele frequencies. Abbreviations are as follows: *EGT* Egypt, *HUT* Hutu (Rwanda), *KEN* Kenya, *TUT* Tutsi, *BHR* Bahrain, *GEO* Georgia, *IRN* Iran, *JOR* Jordan, *OMN* Oman, *PAK* Pakistan, *PUN* Punjabi (India) and *YEM* Yemen. The proportion of variance accounted for by the corresponding scaled data is 95.67%



Partitioning of populations based on geography and language

Distribution of genetic variance along linguistic and geographic lines among the 19 total populations was investigated using hierarchical AMOVA. Potential linguistic correlations were assessed based on the following classification of language families: Niger-Congo (Cabinda of Angola, Hutu, and Tutsi from Rwanda, Kenya and Mozambique); Afro-Asiatic (Bahrain, Egypt, Jordan, Oman, Sudan and Yemen); Kartvelian (Georgia); Indo-European (Belgium, Iran, Pakistan, Punjabis from India and Caucasians from the US); Japanese (Japan); and Sino-Tibetan (Han Chinese from Taiwan). Possible correlations based on geography were ascertained according to the following five regional groups: sub-Saharan Africa (Cabinda of Angola, Hutu and Tutsi from Rwanda, Kenya, Mozambique, and Sudan); Southwest Asia and the Levant region (Bahrain, Egypt, Georgia, Iran, Jordan, Oman and Yemen); South Asia (Pakistan and Punjabi from India); East Asia (Japan and Taiwan); and Europe (Belgium and US Caucasian). Table 3 lists the results locus-by-locus and in combination. The overall linguistic and geographic analyses based on the 15 loci exhibit significant correlation to genetic structure ($P < 0.05$) among groups of populations. With the exception of locus D3S1358 in the linguistic test, all loci individually reflect that the genetic differences among groups of populations correlate significantly ($P < 0.05$) with linguistic and geographical partitioning. On the other hand, the overall linguistic and geographic analyses based on all loci do not significantly correspond to genetic structure among populations within groups. Only three loci (D21S11, TH01, and D18S51) show significant genetic parallelism to language and five (D8S1179, TH01, D13S317, D18S51, and TPOX) to geography.

Discussion

This study presents novel databases for 15 autosomal STR loci from five previously poorly characterized regions (Bahrain, Georgia, Jordan, Kenya, and Sudan). In addition, new data augments previously published autosomal STR datasets for the following six groups: Egypt and Hutu (two novel loci each) (Tahir et al. 2003; Tofanelli et al. 2003, respectively); Oman and Yemen (six new loci each) (Tahir et al. 2000a and Klintschar et al. 2001, respectively); Punjabi (seven novel loci) (Tahir et al. 2000b) and Pakistan (12 new loci) (Hadi et al. 2004). After the application of the Bonferroni adjustment for number of loci tested ($\alpha = 0.05/15$ or 0.0033), only 2% or 3/165 tests fail to conform to Hardy-Weinberg equilibrium expectations (Table 2). More information regarding the intrapopulation STR diversity of these 11 groups is available within the allelic frequency distributions as electronic supplementary material (Tables 1 through 11 at the following URL: http://www.fiu.edu/~herrerar/Frequency_data.htm).

In the NJ analysis (Fig. 2) at the center of the three, the near eastern populations including Sudan, Egypt and the Southwest Asian groups closely encircle the major trifurcation. This intermediary position and tight clustering of these groups may be indicative of their pivotal role of this region in cross continental migrations. It is possible that the dendrogram topology reflects the importance of Southwest Asia and the Levantine region as bidirectional crossroads of human migration involving eastern Asia, Europe and Africa. While this notion is corroborated by the distribution of populations in the MDS plot (Fig. 3), here the affinities of the Arabic and Iranian populations are even more pronounced. It is possible that these populations may have experienced a high degree gene flow not seen in sub-Saharan Africa, East Asia or even South Asia. The phylogenetic analyses

Table 3 Significant locus-by-locus and combined loci AMOVA values for 19 populations

Linguistic		Geographic					
Among groups of populations	Among populations within groups	Among groups of populations	Among populations within groups				
D8S1179	0.80	D21S11	0.28	D8S1179	0.95	D8S1179	0.15
D21S11	1.37	TH01	1.52	D21S11	1.44	TH01	0.81
D7S820	2.66	D18S51	0.12	D7S820	2.62	D13S317	-0.11
CSF1PO	1.09			CSF1PO	1.12	D18S51	0.00
TH01	4.79			D3S1358	0.63	TPOX	0.85
D13S317	3.39			TH01	5.92		
D16S539	1.86			D13S317	3.36		
D2S1338	1.38			D16S539	1.64		
D19S433	1.24			D2S1338	1.45		
vWA	1.84			D19S433	1.22		
TPOX	2.37			vWA	1.85		
D18S51	1.41			TPOX	2.24		
D5S818	2.44			D18S51	1.64		
FGA	0.27			D5S818	2.37		
				FGA	0.41		
<i>Combined</i>	0.10	<i>Combined</i>	NS	<i>Combined</i>	0.10	<i>Combined</i>	NS

Numbers adjacent to locus name refer to percentage of variance significant at $P < 0.05$; NS: Not Significant at $P < 0.05$

Six language groups:

Niger-Congo: Cabinda (Angola); Hutu (Rwanda); Kenya; Mozambique; Tutsi (Rwanda)

Afro-Asiatic: Bahrain; Egypt; Jordan; Oman; Sudan; Yemen

Indo-European: Belgium; Iran; Pakistan; Punjabi; US Caucasian

Kartvelian: Georgia

Japanese: Japan

Sino-Tibetan: Taiwan (Han)

Five geographic groups:

sub-Saharan Africa: Cabinda (Angola); Hutu (Rwanda); Kenya; Mozambique; Sudan; Tutsi (Rwanda)

Southwest Asia & Levant: Bahrain; Egypt; Georgia; Iran; Jordan; Oman; Yemen

South Asia: Pakistan; Punjab

East Asian: Japan; Taiwan (Han)

European: Belgium; US Caucasian

do not indicate any affinity between Arabic and sub-Saharan groups (with the noted exception of the Sudanese population) nor between Arabic and South Asian populations. Overall, the topological layout of the tree and MDS map are consistent with major geographical barriers to gene flow, namely the Sahara in Africa, the Dash-e Kavir and Dasht-e Lut deserts of Iran and the Hindu Kush Mountains on the Afghani-Pakistani border. These obstacles likely have provided an enclosure that allow for gene exchange among populations within its limits while semipermeably encapsulating the region of the Near East.

An examination of the hierarchal AMOVA results (Table 3) indicates that a majority of the loci exhibit significant genetic variance partitioning along both linguistic (14/15 loci) and geographic (15/15 loci) divisions among groups of populations. On the other hand, at most only one-fifth to one-third of the loci exhibit significant correlations to both language and geography, respectively, among populations within groups. It is likely that the allelic distributions at these particular loci provide sufficiently high resolution to detect genetic partitioning along linguistic and geographic lines at the level of populations within groups, which is lost by dilution when combined with the rest of the STR

markers in the overall AMOVA assessment. Since genetic differences are generally larger among groups of populations than within groups of populations, the greater number of loci generating significant correlations to linguistic and geographic partitioning comes as no surprise. The results imply that the genetic structure of the populations in this study parallel both regional subdivision and linguistic hierarchal classification.

Egypt, Arabia and Persia form a tri-continental nexus for initial migrational routes of anatomically modern humans out of Africa and into the remainder of Eurasia beginning some 60,000 YBP (Tishkoff et al. 1996; Watson et al. 1997; Quintana-Murci et al. 1999, 2004) and therefore represent the earliest region of the Asian continent to be inhabited by early man. This conduit was also utilized in “Back to Africa Episodes” (Luis et al. 2004). In addition, the Neolithic advancements in agriculture and animal domestication in the Fertile Crescent about 8,000 YBP likely had a pronounced effect on the contemporary genetic landscape, by nurturing an incubation period of close contact among Near Eastern populations prior to demic diffusion out of the area in all directions. We envision that the Sahara desert to the south and west, the Dash-e Kavir and Dash-e Lut deserts of northern and eastern Iran, respectively, as well

as the Hindu Kush mountains in eastern Afghanistan formed an encirclement of structural barriers to gene flow into and out of the Near East. Since these obstacles do not completely isolate the region, they more likely allowed sporadic, bottleneck-type migrational events (some possibly by coastal routes), thus shaping Arabia and Persia into a basin of genetic homogeneity with narrow passageways of restricted gene flow. In this scenario, it is possible that multiple temporal and spatial bottleneck episodes may account for the decreasing clines of genetic diversity emanating from the Near East eastward into Asia and westward into Europe. Both mtDNA (Quintana-Murci et al. 2004) and Y chromosome studies (Quintana-Murci et al. 2001; Wells et al. 2001; Qamar et al. 2002) suggest the deserts of Persia and the Hindu Kush mountains between the Near East and surrounding territories may have played a key role in limiting gene flow.

Although this study was geared towards a comprehensive characterization of the geographic area from eastern sub-Saharan Africa northeastward into South Asia and the Indian subcontinent, some interesting and specific observations became evident. For instance, within the African clade of the NJ tree the two populations from Rwanda, the Hutu and the Tutsi distinctly separate from each other. Although this phylogenetic analysis indicates a greater differentiation of the Tutsi population in comparison to the other sub-Saharan groups based on branch length in the NJ tree, the bootstrap value is less than 50%. Since both groups cohabit with one another in the same region, it is likely that their genetic uniqueness stems from their socio-political separation (Chretien 2003). It is generally believed that Hutu agriculturalists living in loosely organized clans were the first tribe to supplant the indigenous Twa pygmies of Rwanda. The pastoral Tutsis later settled in from the north, eventually imposing a feudalistic system of government which was the impetus for hundreds of years of social, economic and politically based isolation of these two groups within the same region (July 1992). A maximum likelihood phylogeny in a study based on Y chromosome haplogroup frequencies (Luis et al. 2004) showed that the Hutus segregated closer to a Bantu group of Kenyans than to the Tutsis that segregated into the same clade with mixed Nilo-Saharan and pygmy groups from Central Africa, thus lending support to the evidence of the genetic uniqueness of the Hutus and Tutsis. Yet, the close positioning of the two populations in the MDS analysis suggests some degree of genetic similarity.

Genetically, the Sudanese lie in an intermediary position between the other sub-Saharan populations and the Afro-Asiatic speaking peoples in the phylogenetic results. The fact that the Sudanese segregate closer to the Arab populations than to the sub-Saharan African groups in both NJ and MDS phylogenetic analyses is supported in the literature by studies involving NRY haplogroups in which the frequency of the E3a-M2 mutation characteristic of the Bantu expansion is close

to 0% in the Sudanese (Underhill et al. 2000) and up to 52% in neighboring Kenya (Luis et al. 2004). Bahrain and Oman segregate close to Sudan in both NJ and MDS phylogenetic examinations belying their substantial geographic distance from that African population. This particular arrangement may hint at a possible migrational route and gene flow involving Sudan and the Near East by way of the Nile River waterway and/or the Horn of Africa.

Georgia represents the northernmost border of the Southwest Asian populations examined in this study and displays the longest branch length in the NJ phylogram of that group. The Dash-e Kavir desert in northern Iran may have played a role in the relative singularity of Georgia in comparison to the populations to the south. When a NJ tree was generated using all 19 populations Georgia segregates with the Europeans instead of the Asians (data not shown), which may lend support to this idea. Interestingly, the Iranian group segregates within the Asian clade, but distant from the Pakistani and Punjabi populations, which may argue for a barrier to gene flow between these groups. A recent study of mtDNA on Iranian and Indian populations indicated a similar situation of relatively little gene exchange among the two groups (Metspalu et al. 2004). On the other hand, there is evidence of genetic homogeneity involving Iran and populations to the west within the Near East. This is illustrated in our phylogenetic analyses, which is mirrored in the literature in a report showing genetic homogeneity between Iran and the East Anatolia region of Turkey (Shepard and Herrera 2005). Overall, the results obtained in this study are more compatible with well-defined physical barriers limiting gene flow within the Near East basin and not just the product of differences in geographical distances between populations. This is evident in the distant phylogenetic relationships of geographically close populations like Iran and Pakistan, as well as Yemen and Kenya, in contrast to the strong genetic affinity of geographically distant groups like Sudan and Oman. These affinities are not mere socio-political in nature since regions like Iran and Pakistan share religion, beliefs and culture.

Conclusion

The genetic information generated from this battery of autosomal STR markers not only has applications to population genetics but to recent human evolution as well. The hypervariability of these microsatellites markers generates enough evolutionary signal to clarify genetic relationships of closely related groups. The findings of this study expose a region of genetic homogeneity within the Near East with increasing genetic heterogeneity moving outwards in a southerly, northerly, and easterly direction. The Sahara in Africa, the Dash-e Kavir and Dasht-e Lut deserts of Iran, and the difficult terrain east of Iran in the Hindu Kush Mountain range appear to be the major geographic bounds of

this genetic homogeneity. Substantial geographical confinement would allow for limited gene flow between this region and the outside. Thus, beyond these barriers, we detect a greater degree of genetic differentiation that is consistent with population bottlenecks and more specifically a decrease in genetic diversity in Asia to the east of Persia.

Acknowledgements The authors gratefully acknowledge Dr. Timothy Collins' DNA CORE Facility at Florida International University and Rachel Chow for technical assistance. In addition, we extend our appreciation to Arif Kalantar of the Forensic Laboratory at the Dubai Police Department in the United Arab Emirates and Dr. Marian M. de Pancorbo of la Universidad del Pais Vasco in Spain for their contributions to this study. Lastly, our gratitude goes out to Dr. Gabriel Novick and to Dr. Mohammad Tahir for the invaluable samples from Jordan and Sudan.

References

- ¹URL for linguistic affiliation web resource in Table 1: <http://www.ethnologue.com>
- ²URL for geographical co-ordinate web resource in Table 1: <http://www.nationalgeographic.com/xpeditions/atlas/index.html>
- Alves C, Gusmao L, Damasceno A, Soares B, Amorim A (2004) Contribution for an African autosomal STR database (AmpFISTR Identifiler and Powerplex 16 system) and a report on genotypic variations. *Forensic Sci Int* 139:201–205
- Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C (2004) A predominantly Neolithic origin for Y chromosomal DNA variation in North Africa. *Am J Hum Genet* 75:338–345
- Bar-Yosef O (2002) The upper paleolithic revolution. *Annu Rev Anthropol* 31:363–393
- Beleza S, Alves C, Reis F, Amorim A, Carracedo A, Gusmao L (2004) 17 STR (AmpFISTR Identifiler and Powerplex 16 system) from Cabinda (Angola). *Forensic Sci Int* 141:193–196
- Bosch E, Calafell F, Perez-Lezaun A, Comas D, Mateu E, Bertranpetit J (1997) Population history of north Africa: evidence from classical genetic markers. *Hum Biol* 69:295–311
- Bosch E, Calafell F, Perez-Lezaun A, Clarimon J, Comas D, Mateu E, Martinez-Arias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Hariti G, Cambon-Thomsen A, Bertranpetit J (2000) Genetic structure of north-west Africa revealed by STR analysis. *Eur J Hum Genet* 8:360–366
- Bowcock A, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457
- Brenner C, Morris J (1990) Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. *Proceedings for the International Symposium on Human Identification 1989*. Promega Corporation, Madison, WI, pp 21–53
- Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC (2003) Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American and Hispanic populations. *J Forensic Sci* 48(4):908–911
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The history and geography of human genes*. Princeton University Press, Princeton, NJ
- Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57:133–149
- Chretien JP (2003) *The Great Lakes of Africa: two thousand years of history*. Urzone Inc., New York, NY
- Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 70:1197–1214
- Decorte R, Engelen M, Larno L, Nelissen K, Gilissen A, Cassiman JJ (2004) Belgian population data for 15 STR loci (AmpFISTR SGM Plus and AmpFISTR Profiler PCR amplification kit). *Forensic Sci Int* 139:211–213
- Diamond J, Bellwood P (2003) Farmers and their languages: the first expansions. *Science* 300:597–603
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Felsenstein J (2002) *Phylogeny Inference Package (PHYLIP) Version 3.6a3* distributed by author, Department of Genetics, University of Washington, Seattle, WA
- Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372
- Hadi S, Vanezis P, Goodwin W (2004) The autosomal STR frequencies in Pakistani populations. *J For Sci* 49:868–869
- Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of Human Y chromosome variation. *Mol Biol Evol* 15:427–441
- Hashiyada M, Itakura Y, Nagashima T, Nata M, Funayama M (2003) Polymorphism of 17 STRs by multiplex analysis in Japanese population. *Forensic Sci Int* 133:250–253
- Hassan F (2002) Archeology and linguistic diversity in North Africa. In: Bellwood P, Renfrew C (eds) *Examining the farming/language dispersal hypothesis*. McDonald Institute for Archeological Research, Cambridge, England, pp 127–133
- <http://www.bartleby.com>
- <http://www.nationalgeographic.com/xpeditions/atlas/index.html>
- Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713
- Jones DA (1972) Blood samples: probability of discrimination. *J Forensic Sci Soc* 12:355–359
- Jorde L, Bamsgad M, Watkins W, Zenger R, Fraley A, Krakowiak P, Carpenter K, Soodyall H, Jenkins T, Rogers A (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* 57:523–538
- Jorde LB, Watkins WS, Bamshad MJ, Dizon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal and Y-chromosome data. *Am J Hum Genet* 66:979–988
- July RW (1992) *A history of the African people*, 4th edn. East African Educational Publishers Ltd., Nairobi, Kenya
- Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P, Xiao C, Shriver M, Lell J, Wallace D, Wells RS, Seielstad M, Oefner P, Zhu D, Jin J, Huang W, Chakraborty R, Chen Z, Jin L (2001) African origins of modern humans in East Asia: a tale of 12,000 Y-chromosomes. *Science* 292:1151–1153
- Kivisild T, Bamshad M, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins W, Dixon M, Papiha S, Mastana S, Mir M, Ferak V, Villems R (1999) Deep common ancestry of Indian and Western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331–1334
- Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet* 75:752–770
- Klitschkar M, Al-Hammadi N, Reichenpfeifer B (2001) Significant differences between Yemenite and Egyptian STR profiles and the influence on frequency estimations in Arabs. *Int J Legal Med* 114:211–214

- Levene H (1949) On a matching problem arising in genetics. *Ann Mathm Stat* 20:91–94
- Li CC (1976) First course in population genetics. Boxwood Press, Pacific Grove, CA
- Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: evidence for bi-directional corridors of human migrations. *Am J Hum Genet* 74:532–544
- Lum JK, Jorde LB, Schiefenhovel W (2002) Affinities among Melanians, Micronesians and Polynesians: a neutral biparental genetic perspective. *Hum Biol* 74(3):413–430
- Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13–20
- Maca-Meyer N, Gonzalez AM, Pestano J, Flores C, Larruga JM, Cabrera VM (2003) Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet* 4:15–25
- Manni F, Leonardi P, Barakat A, Rouba H, Heyer E, Klintschar M, McElreavey K, Quintana-Murci L (2002) Y chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Africa. *Hum Biol* 74:645–658
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kalidma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R (2004) Most of the extant mtDNA boundaries in the south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5:26–50
- Militarev A (2002) 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 Archeological Research, Cambridge, England, pp 135–150
- Nebel A, Nadeau-Tasseron E, Filon D, Oppenheim A, Faerman M (2004) Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa. *Am J Hum Genet* 74:532–544
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, NY
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 23:437–441
- Quintana-Murci L, Semino O, Bandelt H, Passarino G, McElreavey K, Santachiara-Benerecetti A (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441
- Quintana-Murci L, Krausz C, Zerjal T, Sayar H, Hammer M, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K (2001) Y chromosome lineages trace diffusion of peoples and languages in Southwestern Asia. *Am J Hum Genet* 68:537–542
- Quintana-Murci L, Chaix R, Spencer Wells R, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Mehdi SQ, Torroni A, McElreavey K (2004) Where west meets east: the complex mtDNA landscape of the Southwest and Central Asian corridor. *Am J Hum Genet* 74:827–845
- Regueiro M (2004): personal communication
- Renfrew C (1996) Language families and the spread of farming. In: Harris D (ed) *The origins and spread of agriculture and pastoralism in Eurasia*. University College of London Press, London, England, pp 70–92
- Renfrew C (2000) At the edge of knowability: towards a prehistory of languages. *Camb Archeol J* 10:7–34
- Rowold DJ, Herrera RJ (2003) Inferring recent human phylogenies using forensic STR technology. *For Sci Int* 133:260–265
- Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diaz P, Macaulay V, Carracedo A (2002) The making of the African mtDNA landscape. *Am J Hum Genet* 71:1082–1111
- Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Schneider S, Kueffer J-M, Roessli D, Excoffier L (2000) *Arlequin v.2000: a software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA (2002) Ethiopians and Khosians share the deepest clades of the human y-chromosome phylogeny. *Am J Hum Genet* 70:265–268
- Shepard EM, Herrera RJ (2005) Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Sci Int* (in press)
- SPSS for Windows, Rel. 11.0.1.2001. Chicago: SPSS Inc
- STRBase (<http://www.cstl.nist.gov/biotech/strbase/>)
- Stringer C (2000) Paleoanthropology: coasting out of Africa. *Nature* 405:24–25, 27
- Tahir MA, Balamurugan K, Tahir UA, Amjad M, Awin MB, Chaudhary OR, Hamby JE, Budowle B, Herrera RJ (2000a) Allelic distribution of nine short tandem repeat (STR), HLA-DQA1 and polymarker loci in an Omani sample population. *Forensic Sci Int* 109:81–85
- Tahir MA, Herrera RJ, Khan AA, Kashyap VK, Duncan G, Barna C, Budowle B, Rowold DJ, Amjad M, Sinha S (2000b) Distribution of HLA-DQA1, polymarker CSF1PO, vWA, TH01, TPOX, D16S539, D7S820, D13S317 and D5S818 alleles in East Bengali and West Punjabi populations from the Indo-Pak subcontinent. *J Forensic Sci* 45:1320–1323
- Tahir MA, Herrera RJ, el-Gohary M, Granoff M, Amjad M (2003) Allele frequency distribution of Power Plex 1.2 and Profiler Plus short tandem repeat (STR) loci in Egyptian population. *J Forensic Sci* 48:889–890
- Tereba A (1999) Tools for analysis of population statistics. Profiles in DNA, v.2, Promega Corporation pp14–16
- Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M (1996) Global patterns of linkage disequilibrium at the CD4 locus in modern human origins. *Science* 271:1380–1387
- Tofaneli S, Boschi I, Bertoneri S, Coia V, Taglioli L, Franceschi MG, Destro-Bisol G, Pascali V, Paoli G (2003) Variation at 16 STR loci in Rwandan (Hutu) and implications on profile frequency estimation in Bantu-speakers. *Int J Legal Med* 117:121–126
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kaufman E, Bonne-Tamir B, Bertrapeit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361
- Wang C-W, Chen D-P, Chen C-Y, Lu S-C, Sun C-F (2003) STR data for the AmpFISTR SGM Plus and Profiler loci from Taiwan. *Forensic Sci Int* 138:119–122
- Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB (2001) Patterns of ancestral human diversity: an analysis of *Alu*-insertions and restriction site polymorphisms. *Am J Hum Genet* 68:738–752
- Watson E, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691–704
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244–10249
- Zvelebil M (1980) The rise of the nomads in Central Asia. In: Sherratt A (ed) *The Cambridge encyclopedia of archeology*. Crown, NY, pp 252–256