

Gene Flow at the Crossroads of Humanity: mtDNA Sequence Diversity and *Alu* Insertion Polymorphism Frequencies in Uzbekistan

Eric J. Devor^{*1}, Ibrokhim Abdurakhmonov², Mark Zlojutro³, Meredith P. Millis¹, Jessica J. Galbraith¹, Michael H. Crawford⁴, Shukhrat Shermetov², Zabardast Buriev² and Abdusattor Abdukarimov²

¹Molecular Genetics and Bioinformatics, Integrated DNA Technologies, Coralville, Iowa 52241 USA

²Laboratory of Genetic Engineering and Biotechnology, Institute of Genetics and Plant Experimental Biology, Academy of Sciences of Uzbekistan, Yuqori Yuz, Tashkent-702151, Uzbekistan

³Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas 78227, USA

⁴Department of Anthropology, University of Kansas, Lawrence, Kansas 66045 USA

Abstract: Mitochondrial DNA (mtDNA) HVS-I region sequences were obtained from 47 unrelated individuals representing 10 of 12 viloyats of Uzbekistan. In addition, frequencies for five *Alu* insertion polymorphisms were determined for the same 47 individuals. These data were used to assess the genetic position of Uzbekistan relative to other Central Asians and, more generally, to Eurasian groups. Results show that the Uzbek sample has an approximate balance of west Eurasian (53.2%) and Asian (46.8%) mtDNA lineages, which is reflected by their intermediate position relative to other Eurasian groups in MDS plots based on genetic distance matrices. The Uzbeks also exhibit high sequence diversity, a fragmented median-joining network, and a low regional Φ_{ST} score, all of which suggests a high degree of gene flow from neighboring gene pools. This is consistent with Central Asia's history of repeated incursions by various nomadic peoples from the Asian steppe and the location of the well-traveled Silk Road within the present-day borders of Uzbekistan. In addition, a star-like cluster within haplogroup H was identified in the Uzbek network, potentially representing a west Eurasian genetic substratum of an earlier occupation by Caucasoid peoples.

Keywords: Mitochondrial DNA, *Alu* polymorphism, genetic history, Uzbekistan.

INTRODUCTION

Central Asia, which encompasses the territories between the Caspian Sea and the western boundaries of present-day China, has been a conduit for cultural, economic, and biological exchanges between Europe and Asia for thousands of years [1]. While much of Central Asia is composed of desolate high altitude regions and broad cold deserts devoid of significant human habitation, its river basins and steppes have been occupied since the Lower Paleolithic [2]. However, in spite of its historical importance during the rise of several far-flung empires and its long record of human habitation, the area is poorly characterized genetically. Several reports have begun to address this deficiency [3-6], revealing that the region has been at the center of widespread demographic processes responsible for molding the present-day genetic variation of local populations, as well as neighboring Eurasian groups. In this study, we contribute to the genetic understanding of Central Asia by reporting mitochondrial DNA sequence diversity and *Alu* element insertion frequencies in Uzbekistan. Our results show comparable frequencies of European and East Asian mtDNA lineages which is reflected by ordination analyses of the pairwise genetic

distance matrices that position the Uzbeks intermediate to western and eastern Eurasian populations, strengthening the view of Uzbekistan as a crossroad of humanity.

SUBJECTS AND METHODS

Population Sample

A panel of 50 DNA samples was prepared from whole blood draws at the Institute of Genetics and Plant Experimental Biology of the Academy of Sciences in Tashkent. The origin of each sample was assigned by the birthplace of the donor's mother. Three samples were eliminated because maternal birthplace was not located within Uzbekistan. The remaining 47 individuals are from ten of the twelve viloyats or political districts of Uzbekistan, with the majority (33 of 47 samples, 70.2%) representing the viloyats of Bukhara, Kashkadarya, and Tashkent (Fig. 1). Thus, the panel provides a broader geographic coverage of Uzbekistan than other samples that are in the literature.

mtDNA Sequencing and RFLP Analyses

A 447 bp mtDNA amplicon containing HVS-I segment of the mitochondrial control region was obtained for each member of the Uzbek panel using the primer set mtFOR: 5'-TCCACCATTAGCACCCAAAGCTA-3' (pos. 15,976 to 15,998; $T_M = 58.4^\circ\text{C}$) and mtREV: 5'-ATTGATTTACGGAGGATGG-3' (pos. 16,422 to 16,402; $T_M = 52.6^\circ\text{C}$). Amplicons were then sequenced in both directions on an Ap-

*Address correspondence to this author at the Molecular Genetics and Bioinformatics, Integrated DNA Technologies, 1710 Commercial Park, Coralville, Iowa 52241, USA; Tel: 319-626-8450; E-mail: rdevor@idtdna.com

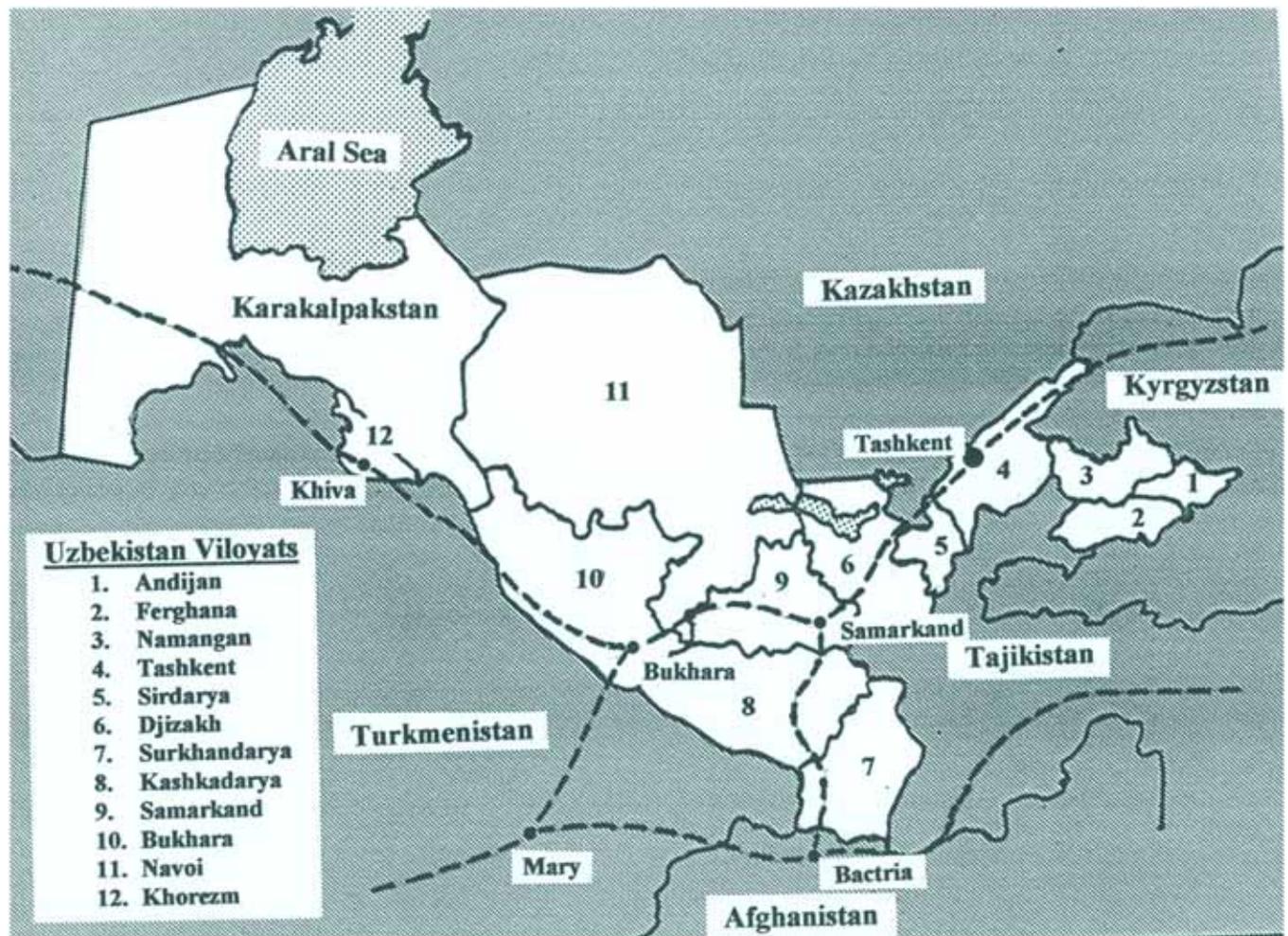


Fig. (1). Map of Uzbekistan. Historic silk routes are shown as dashed lines along with principal regional sites. mtDNA samples were assigned by mother's birthplace to Kashkadarya (13), Tashkent (12), Bukhara (8), Surkhandarya (4), Namagan (3), Samarkand (2), Khorezm (2), Djizakh (1), Andijan (1), and Sirdarya (1).

plied Biosystems Model 310 automated fluorescence sequencer.

Additional characterization of mtDNA haplogroup assignments was obtained, as needed, by RFLP typing. The 10,397 *Alu I* and 10,394 *Dde I* cut sites were typed by restriction enzyme digestion of a 269 bp amplicon produced by the primers 10,290 FOR: 5'-GCCCTACAAACAATAACCTGCCA-3' ($T_M = 59.2^\circ\text{C}$) and 10,558 REV: 5'-AGGGAGGATATGAGGTGTGAGCGAATA-3' ($T_M = 60.1^\circ\text{C}$). The 12,406 *Hinc II/Hpa I* cut site was typed by restriction enzyme digestion of a 215 bp amplicon produced by primers 12,361 FOR: 5'-ACCGAGAAAGCTCACAAGAAGTGC-3' ($T_M = 59.2^\circ\text{C}$) and 12,575 REV: 5'-AGCTTAAGGGAGAGCTGGGTTGTTT-3' ($T_M = 60.0^\circ\text{C}$). These primer sets were designed and analyzed for melting temperature and secondary structure using the on-line SciTools software at Integrated DNA Technologies (www.idtdna.com). PCR conditions were $94^\circ\text{C}^{5:00}$ [$94^\circ\text{C}^{0:30}$; $53^\circ\text{C}^{0:30}$; $72^\circ\text{C}^{0:30}$] $_{35}$ $72^\circ\text{C}^{7:00}$ for the HVR I amplifications and $94^\circ\text{C}^{5:00}$ [$94^\circ\text{C}^{0:30}$; $58^\circ\text{C}^{0:30}$; $72^\circ\text{C}^{0:30}$] $_{35}$ $72^\circ\text{C}^{7:00}$ for the RFLP amplifications. Restriction enzyme digests of the RFLP amplicons were carried out as per supplier recommendations.

Alu Insertion Polymorphisms

Alu elements are a class of retroposons unique to primate genomes [7]. There are more than one million *Alu* elements in the human genome [8] and a few are young enough that they have yet to become fixed. Polymorphic *Alu* insertions are two allele loci easily typed by PCR amplification. We chose five loci for typing the Uzbek panel; FXIIIIB, TPA25, APO, PV92, and ACE. Summary information on each locus, including PCR primer sequences and PCR reaction conditions, are presented in Table 1.

Comparison Population Data

Comparative information for mtDNA sequences and *Alu* insertion polymorphism frequencies was assembled for a number of populations both within and outside Central Asia. MtDNA sequences were obtained for the Central Asian populations Kazakhs, Kirghiz, and Uighurs [3]. Southern Siberian groups are represented by the Altai [9], Tuva and Buryats [10], while Eastern Asia is represented by the Mongols [11] and Han Chinese [12]. European groups consist of Russian [13], Greek [14], and Anatolian Turk [15] samples.

Table 1. Primer Sequences and PCR Conditions for the *Alu* Insertion Polymorphisms

Alu ID	Chr.	5' Primer Sequence 3'	Tm(°C) ^a	Ta(°C) ^b	Alleles ^c
FXIII B	1	TCAACTCCATGAGATTTTCAGAAAGT CTGGAAAAAATGTATTCAGGTGAGT	56.2 55.3	54.0	500;200
TPA25	8	GTAAGAGTTCCGTAACAGGACAGCT CCCCACCCTAGGAGAACTTCTCTTT	59.8 61.3	58.0	424;104
APO	11	AAGTGCTGTAGGCCATTTAGATTAG AGTCTTCGATGACAGCGTATACAGA	56.5 58.9	55.0	409;83
PV92	16	AACTGGGAAAAATTTGAAGAGAAAAGT TGAGTTCTCAACTCCTGTGTGTTAG	55.3 58.0	55.0	442;119
ACE	17	CTGGAGACCACTCCCATCCTTTCT GATGTGGCCATCACATTCGTCAGAT	61.6 61.0	58.0	470;173

^aSequence melting temperature estimated by OligoAnalyzer 3.0 (see SciTools at www.idtdna.com).

^bTemperature set for primer annealing step of the PCR assay. Constant conditions are: 94°C^{5:00}[94°C^{0:30}; Ta(°C)^{0:30}; 72°C^{0:30}]₃₅72°C^{7:00}.

^cAmplicon length with and without the *Alu* insertion.

FXIII B, TPA25, APO, PV92, and ACE *Alu* insertion frequencies used for population comparison were obtained from the ALFRED database at Yale University (<http://alfred.med.yale.edu/alfred/index.asp>), including Uighur, Tajik, Buryat, Turk and Greek populations. In addition, several Asian population samples were typed in the present study. These include the Yakuts (n = 50) and Mongols (n = 50), as well as the Chuvash (n = 50) and Koryaks (n = 19) that have not been previously typed for these particular markers. Polymorphism frequencies from these additional samples have been deposited in ALFRED.

Data Analysis

Using the software package Arlequin ver. 2.0 [16], three different sequence diversity measures were employed: (i) haplotypic diversity $[(n/n-1)(1-\sum p_i^2)]$, where p_i is the frequency of the i th haplotype; (ii) θ_π estimator of the parameter θ ($2N_e\mu$), which is equivalent to the mean pairwise differences between sequences (π); and (iii) θ_S estimator that is based upon the number of observed nucleotide variant sites (S) within a sequence data set. AMOVA analyses were performed for the entire HVS-I data set, as well as for defined regional groups (see below). The neutrality test statistics Tajima's D [17] and Fu's F_S [18] and the τ parameter of Rogers and Harpending's mismatch model [19] were also computed for the Uzbek sample.

A median-joining (MJ) network was constructed for the Uzbek HVS-I sequences using Network ver. 4.0 [20, 21]. The MJ algorithm generates the most parsimonious trees of observed haplotypes by resolving parallelisms with a compatibility argument justified by multi-state frequency criteria. Unresolvable parallelisms are highlighted as reticulations in a single diagram.

Nei's intermatch-mismatch (D_A) genetic distances [22] between the population data sets were computed using Arlequin ver. 2.0. $D_A = d_{ij} - (d_{ii} + d_{jj})/2$, where d_{ij} is the intermatch mean (i.e., the mean number of nucleotide differences) between populations i and j , and d_{ii} and d_{jj} are the mismatch values within populations i and j , respectively. Since a tree representation of genetic distances may be misinterpreted as a succession of population splits, the D_A matrix was displayed in two-dimensional space by multidimen-

sional scaling (MDS) using the NTSYS statistical package [23].

For the *Alu* polymorphism frequencies, the coefficient of gene differentiation, $G_{ST} = (H_T - H_S)/H_T$, was computed for each locus using the executable DISPAN [24], where H_T is the gene diversity for the total population (i.e., average allele frequencies for the entire data set) and H_S is the average of the gene diversities computed for the individual (sub)populations [25]. Nei's modified Cavalli-Sforza genetic distance [26] was calculated between the populations and visualized using the MDS methodology.

Results

mtDNA Sequence Diversity

mtDNA sequences obtained from the Uzbek panel are presented in Fig. (2) and summary statistics for the Uzbek and comparison populations are shown in Table 2. A total of 44 different mtDNA sequences are observed in the Uzbek panel. These sequences account for 57 variable positions, of which 56 represent transition events and a lone transversion occurring at np16183 (A→C). All three of the population diversity measures based upon mtDNA sequences (Table 2) clearly show that the Uzbek sample is highly diverse and similar to the levels exhibited by the other Central Asian populations. The South Siberians, on the other hand, are collectively less diverse than the other Asian populations listed in Table 2, suggesting a distinct demographic profile for the history of this particular region. The European groups display diversity scores that generally fall between the range of values for the Central and East Asians and the South Siberians. The extremes of sequence diversity, at least within this limited comparison, are represented by the Buryats and Han Chinese (with the exception of the Altai who have the smallest θ_S value). In the former, a substantial proportion of the HVS-I data set are characterized by a single HVS-I sequence (25.0%) whereas the latter population sample is dominated by low-frequency haplotypes with varied geographic distributions.

Median-Joining Network and Neutrality Test Scores

Based on the network analysis of the HVS-I sequences (Fig. 3), the Uzbek sample was assigned to a large number of

Table 2. Population Diversity Measures Based Upon mtDNA Sequences for the Uzbek and Comparison Populations

Population (n)	Haplotypic Diversity (sd) ^a	$\theta\pi$ (sd) ^b	θ_s (sd) ^c	Source
Central Asia				
Uzbek (47)	0.996 (0.005)	5.919 (3.193)	12.906 (3.936)	This Study
Kazakh (55)	0.990 (0.006)	6.611 (3.517)	13.988 (4.116)	[3]
Kirghiz (95)	0.990 (0.004)	6.276 (3.329)	15.640 (4.154)	[3]
Uighur (55)	0.993 (0.005)	5.848 (3.148)	13.769 (4.057)	[3]
South Siberia				
Altai (17)	0.993 (0.023)	5.522 (3.126)	7.986 (3.137)	[9]
Tuva (36)	0.978 (0.015)	6.887 (3.678)	10.128 (3.324)	[10]
Buryat (40)	0.936 (0.031)	3.967 (2.252)	9.169 (2.981)	[10]
East Asia				
Mongol (103)	0.995 (0.002)	6.474 (3.420)	14.788 (3.888)	[11]
Han Chinese (250)	0.999 (0.001)	7.675 (3.972)	20.995 (4.683)	[12]
Non-Asian				
Russian (103)	0.965 (0.011)	4.222 (2.340)	11.331 (3.059)	[13]
Anatolian Turk (45)	0.993 (0.007)	5.343 (2.916)	12.349 (3.812)	[15]
Greek (54)	0.990 (0.006)	4.533 (4.533)	10.972 (3.313)	[14]

^aComputed based upon haplotype numbers (see [16]).

^bComputed based upon mean pairwise distances.

^cComputed based upon number of variant sites.

different mtDNA lineages. Within macrohaplogroup M (defined by restriction cut sites at 10394 *DdeI* and 10397 *AluI*), the following (sub)haplogroups are observed: C, D, G2a, M7c, M10, M*, and Z. For macrohaplogroup N, which represents the other major mtDNA phylogenetic branch for Eurasian populations, the following are present: A, B4, F, H, J, N*, T, U4, and U5. This spectrum of haplogroups is consistent with those reported for other Central Asian populations [5, 6].

The overall topology of the network is highly fragmented, displaying long branches between observed haplotypes, with the lone exception being the star-like component centered on haplogroup H. This type of population genetic structure is often associated with population bottlenecks or sustained gene flow from both neighboring and distant gene pools. The neutrality test statistics Tajima's *D* and Fu's *F_S* are both negative (-1.898 and -25.293, respectively) and significant at the 0.01 level, indicating an inflation in the number of variant sites (*S*) relative to the mean pairwise differences (π) and based on coalescent theory is typically interpreted as evidence for population expansion [27, 28]. However, as stated above, the fragmented network structure shows little evidence for this. According to a simulation study conducted by Ray *et al.* [29], gene flow (*Nm*) was shown to have an impact on both neutrality measures, with large *Nm* values generating highly significant negative test scores through the introduction of low-frequency alleles. Thus, the Uzbek negative scores may be interpreted as a product of the sample's high diversity and the large number of low-frequency, phylogenetically isolated haplotypes introduced into the population by regional gene flow (i.e., 40 of the 44 different sequences have a frequency of one, with

55% of these removed from other haplotypic nodes in the network by at least three variant HVS-I sites). Furthermore, the negative scores repudiate any genetic impact of population bottlenecks, since low-frequency haplotypes would likely be purged in such an event, increasing intermediate- and high-frequency variants that would inflate the value of π relative to *S* and thus theoretically pushing both neutrality test statistics towards positive scores.

Among the comparative populations, the Central Asian groups, Buryats, Mongols and Han Chinese all exhibit significant negative scores ($P < 0.05$) for both Tajima's *D* and Fu's *F_S* (not shown). Similar to the Uzbeks, the Central Asians have fragmented network topologies, whereas the Mongol and Han samples both exhibit elaborate star-like features within various phylogenetic branches, including haplogroup D [12, 30], the main mtDNA lineage for both populations (24.3 % and 21.2%, respectively). Likewise, the European groups in this study have highly significant scores ($P < 0.01$) and star-like genealogies, which are common genetic signatures for Europeans that have been largely attributed to the Paleolithic expansion of humankind into the continent and the dispersion of Middle Eastern farmers during the Neolithic period [31, 32]. Therefore, in relation to East Asia and Europe, Central Asia appears to be a zone of significant gene flow, with its high sequence diversity, fragmented genealogies, and significant negative neutrality test scores.

Mismatch Analysis

The mismatch distribution for the Uzbeks (Fig. 4) is unimodal, with a peak at 5-6 pairwise differences, which according to coalescent theory (19) indicates past population

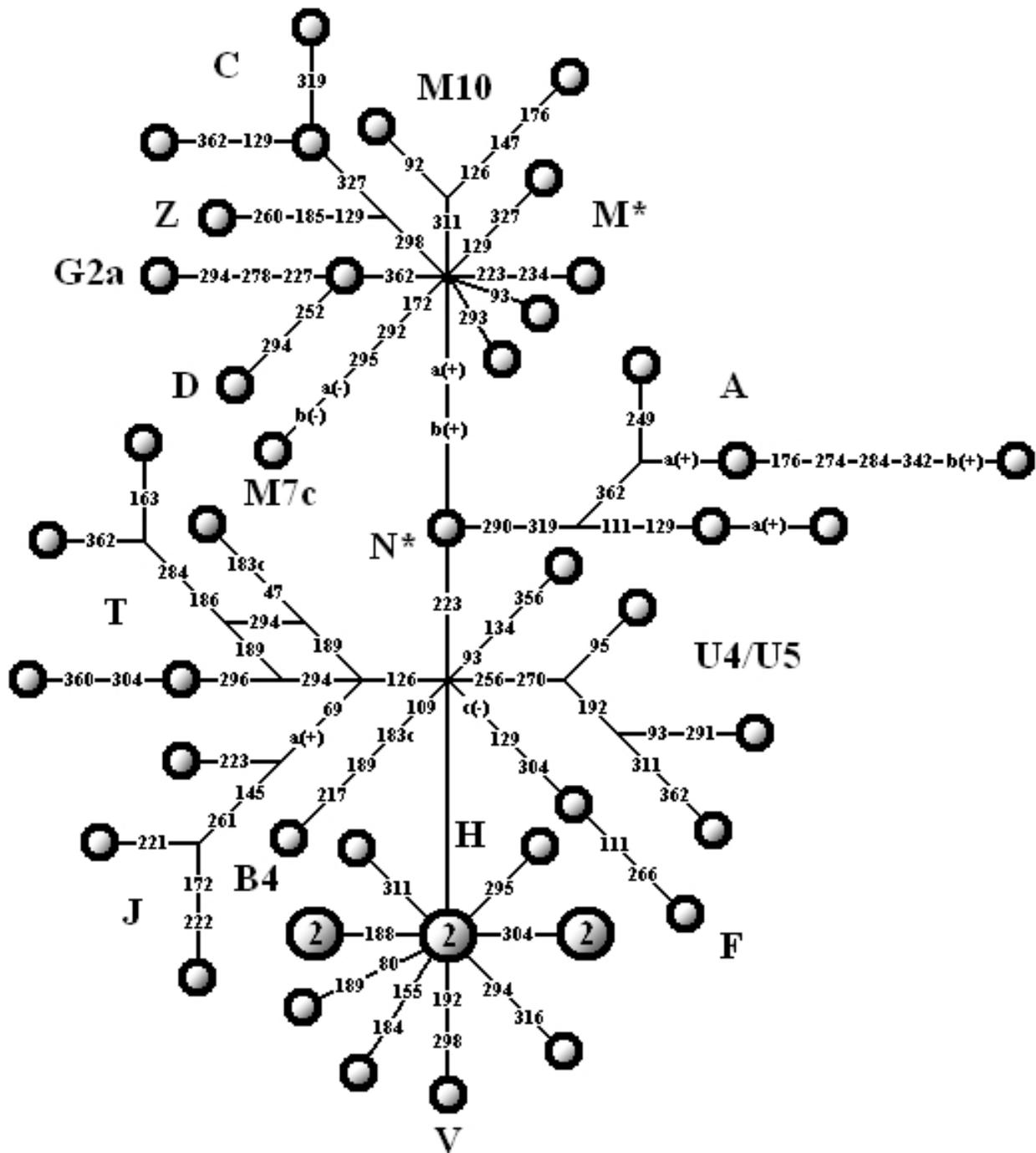


Fig. (3). Median-joining (MJ) network of the Uzbekistan mtDNAs. The HVSI transitions are abbreviated to the last three digits of the nucleotide position (np 16000-16400) and are relative to the CRS sequence, with the lone transversion, 16183A→C, listed as 183c. Diagnostic RFLPs were used to resolve some of the haplogroup assignments. These RFLPs are represented by the following letters: a - 10397 *Alu I*; b - 10394 *Dde I*; and c - 12406 *Hpa I/Hinc II*. The presence (+) or absence (-) of a cut site is indicated in parentheses.

growth. The coalescent date derived from the mismatch distribution using the fast and more conservative HVS-I mutation rate of 16.5% per nucleotide site per million years [33] is approximately 50,700 YBP (95% CI: 32,300-61,400 YBP). This date is remarkably close to the times estimated for the other Central Asian populations (49,400-51,800 YBP). Given this time estimate, the unimodality of the mismatch distribution is likely to be a signature of ancient demographic processes associated with the initial expansion

of humankind into Central Asia during the Paleolithic period, which is consistent with previous mismatch analyses of Eurasian populations [34, 35].

European-Asian Sequence Sharing

Diversity statistics for Uzbek mtDNAs, while useful, do not address the major issue of assessing the structure of diversity at regional levels. A sequence by sequence comparison between the populations (not shown) indicate that four-

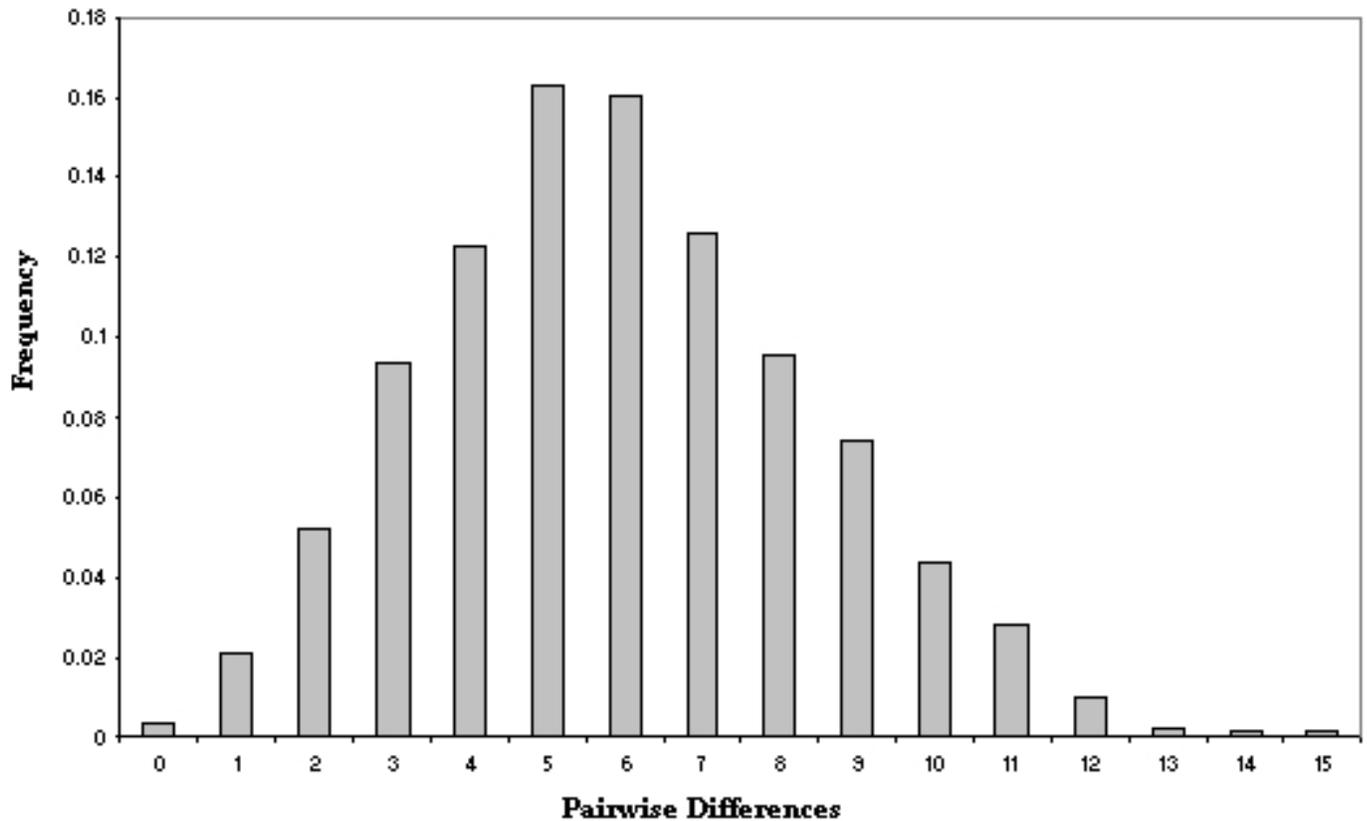


Fig. (4). Mismatch distribution for the Uzbek sample.

teen of the 44 mtDNA sequences observed in the Uzbek sample are shared with the Mongols, four with the Han, six with both the Russians and Turks, and three with the Greeks. Among the shared Uzbek sequences, those corresponding to traditionally European haplogroups H, T, J, and U are predominantly seen in the Central Asian and European samples, while those assigned to traditionally Asian haplogroups A, C, and D are most frequent in the Central Asian, Siberian, and East Asian samples.

In a similar vein, Comas *et al.* [5] classified their Central Asian sequences, which included a small sample of Uzbeks ($n=20$), to either a west Eurasian, East Asian, or South Asian/Indian mtDNA category based on diagnostic variant sites and phylogeographic information from previously reported Eurasian population data sets. The authors assigned 45.8% of their total Central Asian sample to west Eurasian mtDNA lineages, 51.6% to East Asian, and 2.6% to South Asian (defined by haplogroups M*, M2 – M6) [36]. Based on this scheme, the Uzbek subset was defined as 50% west Eurasian, 45% East Asian and 5% South Asian. Quintana-Murci *et al.* [6] also presented HVS-I data for a circumscribed Uzbek sample collected from the southernmost viloyat of Surkhandarya ($n=42$), with the following composition: 52.4% west Eurasian, 30.9% East Asian, and 16.7% South Asian.

In the present study, the Uzbek sequences represent far greater coverage of Uzbekistan than the aforementioned papers and can be defined as 53.2% west Eurasian (haplogroups H, J, T, U4, U5 and V), 38.3% East Asian (haplogroups A, B4, C, D, F, G2a, M7c, M10, and Z), and 8.5% South Asian (haplogroup M*). Despite the geographic

differences between the three sample sets, the frequencies are similar and show an approximate balance between west Eurasian and Asian mtDNA lineages. For the comparative populations, a continent-wide gradient is apparent. To the west, the Greek, Russian, and Anatolian Turk samples can be sorted into mostly European and Middle Eastern groupings with few Asian lineages represented. Conversely, the South Siberian and East Asian groups exhibit sequences that predominantly belong to Asian mtDNA lineages (>80%), with a large percentage identified to haplogroups C and D (which collectively range from 24.4% – 52.8%). Thus, the Uzbeks appear to occupy a central position with respect to the mtDNA profile of Eurasian populations.

AMOVA

The genetic structure of the HVS-I data set was further investigated by AMOVA analyses. When the 12 comparative populations were treated as a single group, the proportion of variation among populations (Φ_{ST}) is 0.0456, which is highly significant at the 0.001 level. When the populations are placed into regional groups (as defined in Table 2), the Φ_{ST} is 0.0525, with the proportion of variation among the groups (Φ_{CT}) estimated at 0.03971 (for both $P<0.001$). Analyzing the Central Asian populations separately, the Φ_{ST} is 0.0053 ($P=0.055$), which contrasts the Φ_{ST} values for the other regions (South Siberia, East Asia and Europe) that exhibit values that are about 2-8 times larger (for each $P<0.001$). And when the Central Asians are excluded from the data set, the Φ_{ST} increases from 0.0456 to 0.0655 ($P<0.001$). These AMOVA results collectively show that the Central Asian populations exhibit less differentiation relative to other re-

gional groups, and suggest an important role for the homogenizing effects of gene flow within Central Asia.

MDS Plot of D_A Distances

An intermatch-mismatch (D_A) distance matrix based upon all twelve comparative populations was generated and a multidimensional scaling (MDS) representation of this matrix is shown in Fig. (5). The stress value for the plot is 0.044, which, according to Kruskal [37], indicates an excellent goodness of fit to the original distance matrix. The first axis displays an East-West orientation that is generally consistent with the geographic locations of the populations, while the second axis primarily delineates the Han Chinese from the other Asian populations. On both axes the Uzbek population holds a central position. From the perspective of the three mtDNA categories described above, the first axis is predominantly influenced by the proportion of population samples that are classified to European or Asian lineages, whereas the second axis reflects significant haplogroup differences among the Asian groups themselves (e.g., haplogroup B has low frequencies in the Asian populations used in this study, with the exception of the Han sample with 16.8% belonging to this haplogroup). When considering linguistic affiliation, the Turkic-speaking peoples maintain the East-West geographic orientation, with the exception of the Altai who are removed from this group relative to the second axis.

Alu Insertion Polymorphisms

Allele frequencies for the five *Alu* insertion polymorphisms are shown in Table 3. With the exception of APO, the total gene diversity (H_T) among the populations is quite high, ranging from 0.373-0.485. However, most of the gene diversity is attributable to diversity between individuals within populations (H_S). The degree of genetic differentia-

tion among populations (G_{ST}) ranges from 0.030-0.160, with an average of 0.0878. A multidimensional scaling (MDS) representation of genetic distances based upon these allele frequencies is shown in Fig. (6). The stress value is 0.0373, which indicates a very robust fit with the original distance matrix. Again, as with the mtDNA sequences, the first axis is driven by European-Asian differentiation while the second axis is driven by differentiation within Asia itself. Also, as seen with the mtDNA sequences, Uzbekistan occupies a central position on both axes which holds when the non-Turkic speaking populations are removed from the array.

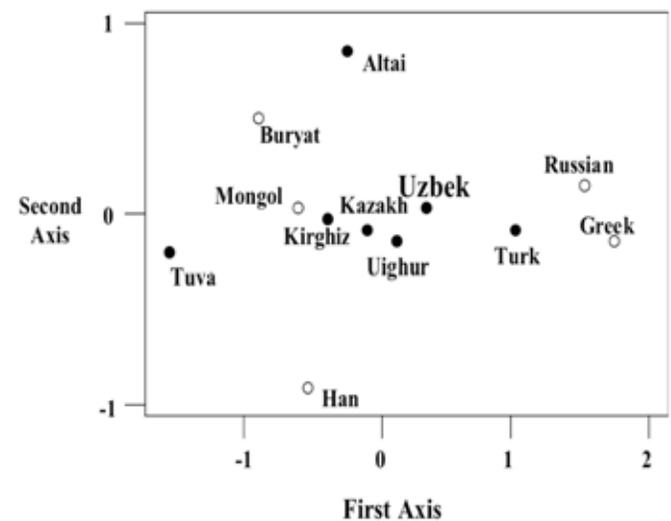


Fig. (5). Multidimensional scaling (MDS) representation of the mtDNA intermatch-mismatch distance matrix for Uzbek and eleven comparison populations. Turkic-speaking groups are indicated by closed circles and non-Turkic-speaking groups are indicated by open circles.

Table 3. Allele Frequencies for the Five *Alu* Insertion Polymorphisms Among the Uzbek and Comparison Populations

Population (n)	FXIII B	TPA25	APO	PV92	ACE	Source ^a
Uzbek (47)	0.494	0.505	0.508	0.508	0.115	This Study
Uighur (85)	0.347	0.504	0.096	0.501	0.496	ALFRED
Tajik (34)	0.502	0.515	0.202	0.376	0.510	ALFRED
Chuvash (50)	0.486	0.508	0.167	0.437	0.507	This Study
Mongol (50)	0.361	0.429	0.231	0.464	0.508	This Study
Buryat (46)	0.261	0.510	0.150	0.511	0.403	ALFRED
Yakut (50)	0.288	0.503	0.327	0.339	0.458	This Study
Han Chinese (49)	0.420	0.464	0.000	0.246	0.451	ALFRED
Koryak (19)	0.061	0.514	0.100	0.514	0.284	This Study
Turk-Cypriot (33)	0.491	0.502	0.040	0.456	0.456	ALFRED
Greek-Cypriot (50)	0.481	0.508	0.097	0.383	0.486	ALFRED
H_S^b	0.373	0.485	0.162	0.420	0.450	
H_T^c	0.431	0.500	0.168	0.500	0.488	
G_{ST}^d	0.136	0.030	0.035	0.160	0.078	

^aAllele frequencies determined in groups from this study not previously listed in ALFRED have been deposited under standard submission criteria.

^bwithin subpopulations.

^cbetween subpopulations.

^dCoefficient of gene differentiation (see [25]).

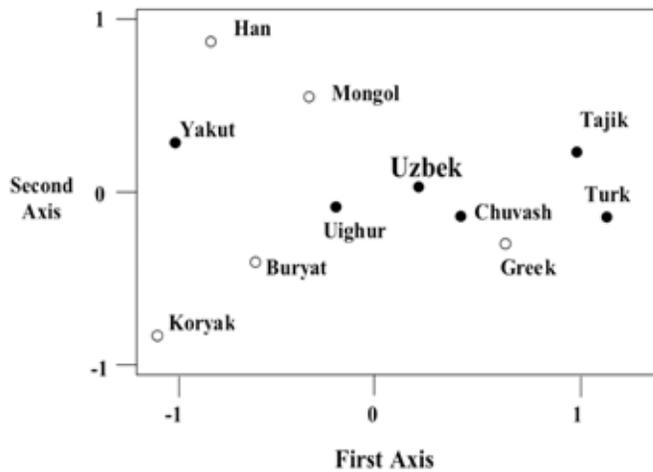


Fig. (6). Multidimensional scaling (MDS) representation of the genetic distance matrix for *Alu* insertion polymorphism frequencies from Uzbek and ten comparison populations. Turkic-speaking groups are indicated by closed circles and non-Turkic-speaking groups are indicated by open circles.

DISCUSSION

Geographic, cultural, and historical factors have combined to make Central Asia an active nexus for human migration within the Eurasian continent. Starting around 5,000 YBP, the domestication of the horse, followed by the development of wheeled vehicles, gave the inhabitants of the Central Asian steppes a means to expand geographically [38, 39]. These innovations spurred repeated population waves in this region until the Middle Ages, first of Caucasoid and Indo-European speaking peoples, then by the sixth century BC of East Asian nomads speaking languages belonging to the Altaic family [40]. Early Caucasoid migrations from the Andronovo and Srubnaya cultures of the Volga steppe are believed to be responsible for establishing the vast Central Asian settlements attributed to the pastoral Scythians and, later, Sarmatians. By the sixth century BC, the Asian steppes were invaded and subsequently ruled by Mongol nomads known as the Hsiung-nu for several centuries [41]. Shortly after their disappearance from historical chronicles, the T'u-chüeh (or Kök-Turks) from Inner Mongolia replaced the Hsiung-nu as the most powerful nomads of Central Asia during the sixth century AD, and thus initiating a historical trend of Mongol-Turkic domination that reached its ultimate manifestation with Chingis Khan's continent-wide empire during the 12th and 13th centuries AD.

In addition to the repeated incursions of various nomadic groups, the well-traveled Silk Road has also contributed to the movement of humanity and trade goods across Central Asia since its establishment in the second century BC by the Ch'in dynasty. Uzbekistan encompasses the main arteries of traffic beyond the Tarim basin and central plains of China, with four of the major historical trade centers on the Silk Road – Khiva, Bukhara, Samarkand, and Tashkent – lying within its present-day borders, and a fifth center, Bactria, located just to the south in Afghanistan (Fig. 1). Genetic results from both mtDNA sequences and five nuclear DNA *Alu* insertion polymorphisms are consistent with this historical narrative, revealing a genetic imprint of the movement of

armies, trade goods, culture, and language through Uzbekistan from both east and west directions.

Several models for the genetic population history of Central Asia have been proposed [42]. From recent genetic studies, three scenarios have been evaluated [5, 6]. First, that Central Asia was settled approximately 40,000 to 50,000 years ago and represents the region where non-African mtDNA matured and differentiated into major Eurasian haplogroups prior to human expansion into Europe and East Asia. Alternatively, the second and third scenarios both posit that Central Asia served as a thruway in the colonization of Eurasia during the Paleolithic period and that more recent history marked by generations of gene flow is primarily responsible for molding the contemporary genetic structure of the region. Specifically, the second scenario holds that Central Asia was as a “contact zone” between differentiated groups of peoples from east and west Eurasia, whereas the third scenario posits that Central Asia was a “hybrid zone” for its original inhabitants, western Caucasoid peoples, who were partially replaced by eastern peoples. The prevailing conclusion among the few genetic studies of extant Central Asian populations favors the second scenario though there is evidence that some putatively ancient Central Asian-specific lineages persist [5].

Our results, specifically the mtDNA lineage composition of the Uzbeks and MDS reductions of genetic distances based on HVS-I sequences and *Alu* polymorphisms suggest that Uzbekistan is the genetic fulcrum of Central Asia. The Uzbek sample displays an approximate balance between west Eurasian and Asian mtDNA lineages, for which each of these mtDNA groupings exhibits opposing east-west frequency gradients among the other comparative population samples included in this study. This pattern is also reflected in the MDS plots, which place the Uzbeks intermediate to the European and Asian groups in two-dimensional space. Therefore, given the lack of substantial overlap between present-day western and eastern Eurasian haplogroups, the first scenario appears to be unlikely. The genetic patterns in and of themselves do not conclusively favor one of the two remaining scenarios, but the MJ network and the AMOVA results appear to more consistent with the notion of Central Asia as a contact zone between eastern and western peoples. The network has an overall shattered topology consisting of long internal branches void of expressed haplotypic nodes, which, given the large diversity of both western and eastern Eurasian mtDNA lineages, is likely to be the product of gene flow rather than autochthonous population expansion that is typically associated with star-like clusters. Furthermore, the Φ_{ST} value for the Central Asians is smaller relative to the other regions considered in this study and non-significant at the 0.05 level, suggesting an important role for gene flow between the populations of Central Asia.

However, despite the overall fragmented structure of the Uzbek network, a lone star-like cluster is evident within haplogroup H. This mtDNA lineage has its largest frequencies in European populations, for which it also exhibits a star-like phylogeny [43]. A coalescent age of 19,200-21,400 YBP (95% confidence interval) for haplogroup H in Europeans was estimated by Richards *et al.* [31] using the ρ statistic and a mutation rate of one transition per 20,180 years [44]. Using the same methodology (Network ver. 4.0), the age of

haplogroup H for the Uzbek sample is approximately 22,000 YBP (ρ statistic is 1.091 ± 0.364), which is close to the European values and younger than the 95% CI ages derived for Near Eastern populations (24,300-29,000 YBP). Furthermore, sequence sharing for Uzbek haplogroup H (10 sequences, 21.3%) was observed exclusively with the Russian, Greek and Turk data sets analyzed in this study, which suggests that the star-like component may represent an earlier occupation of the region by European populations. This scenario is consistent with several facts that point to the presence of western peoples in Central Asia: ancient Chinese texts and the Greek historian Herodotus describe the Scythians as having European morphological traits [45]; extinct Indo-European language, Tocharian, was spoken in the area during the latter half of the first millennium AD [46]; and the presence of mummified bodies with European facial traits in Xinjiang province, China, such as blond or reddish hair and straight noses [47]. Ancient mtDNA analyses also support this scenario. Lalueza-Fox *et al.* [48] typed 27 skeletal remains from Kazakhstan dating between the 15th century BC and 5th century AD for the HVS-I and diagnostic RFLPs in the coding region. The authors discovered a west Eurasian genetic substratum in their sample. Specimens dating prior to the 7th century BC ($n = 13$) were all characterized to west Eurasian haplogroups, with four samples from this subset representing haplogroup H. After this period, 50% of the remaining samples ($n = 14$) were typed to west Eurasian lineages, 42.9% to East Asian, and 7.1% to South Asian. It is noteworthy that the westward movement of Asian nomads, namely the Hsiung-nu, is estimated to have taken place around the 6th century BC, and thus this historical development shows close correspondence to the genetic discontinuity. Remarkably, the mtDNA frequency values of the post-7th century BC subset are very close to those reported for the Uzbeks in the present study (53.2%, 38.3%, and 8.5%, respectively). Three of the Uzbek HVS-I sequences, two belonging to haplogroup H and one to haplogroup T, are shared with this ancient mtDNA data set. Therefore, when considering the historical, archaeological and genetic evidence outlined above, the hybrid zone scenario, which posits early occupation by western Caucasoid peoples followed by East Asian admixture, appears to be the most parsimonious model for the genetic history of Central Asia.

On a final note, Comas *et al.* [5] identified two putatively ancient Central Asian mtDNA lineages, G2a (16223T, 16227G, 16278T, 16362C) and D4c (16245T). Interestingly, only G2a was identified in the Uzbeks, observed in an individual from Surkhandarya, the southernmost of Uzbekistan's viloyats. The other mtDNA lineage, which was reported at appreciable frequencies in other Central Asian populations, is not evident in either our sample nor among Lalueza-Fox *et al.*'s ancient HVS-I sequences, and thus weakening the argument that D4c represents an ancient lineage of Central Asia.

Overall, the genetic character of our sample, a broader representation of the Uzbek population than those of other studies, places Uzbekistan at the center of ancient and widespread demographic processes that have been largely responsible for the contemporary genetic structure of Asian populations. However, additional genetic research still needs to be done on this region. Initial results from Y-chromosomal markers for Central Asian groups conflict with the genetic

picture produced by mtDNA by showing a high degree of genetic differentiation between populations [4], a global trend that has been attributed to differential male and female migration patterns [49]. Furthermore, the presence of a west Eurasian genetic substratum, or lack thereof, needs to be more fully investigated using other genetic systems. As further details of the genetics of Uzbekistan emerge and are considered within the context of its history and culture, its unique role as a crossroads of humanity will undoubtedly expand.

REFERENCES

- [1] Grousset R: *Empire of the Steppes* translated from French by Walford N. New Jersey, USA: Rutgers University Press; 1970.
- [2] Davis RS, Randov VA. Recent work on the Paleolithic of Central Asia. *Evol Anthropol* 1999; 8: 186-93.
- [3] Comas D, Calafel F, Mateu E, *et al.* Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 1996; 63: 1824-38.
- [4] Perez-Lezaun, A, Calafel F, Comas D, *et al.* Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am J Hum Genet* 1999; 65: 208-19.
- [5] Comas D, Plaza S, Wells RS, *et al.* Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet* 2004; 12: 495-504.
- [6] Quintana-Murci L, Chaix R, Wells RS, *et al.* Where West meets East: the complex mtDNA landscape of the Southwest and Central Asian corridor. *Am J Hum Genet* 2004; 74: 827-45.
- [7] Kapitonov V, Jurka J. The age of Alu subfamilies. *J Mol Evol* 1996; 42: 59-65.
- [8] Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet* 2002; 3: 370-9.
- [9] Shields GF, Schmiechen Am, Frazier BL, *et al.* MtDNA sequences suggest a recent evolutionary divergence for Beringian and Northern American populations. *Am J Hum Genet* 1993; 53: 549-62.
- [10] Derenko MV, Malyarchuk BA, Dambueva IK, *et al.* Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of North Asia. *Hum Biol* 2000; 72: 945-73.
- [11] Kolman CJ, Sambuughin AM, Bermingham E. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 1996; 142: 1321-34.
- [12] Yao YG, Kong QP, Bandelt HJ, *et al.* Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 2002; 70: 635-51.
- [13] Orekhov V, Poltoraus A, Zhivotovsky LA, *et al.* Mitochondrial DNA sequence diversity in Russians. *FEBS Lett* 1999; 445: 197-201.
- [14] Kouvatzi A, KaraiskouN, Apostolidis A, Kirmizidis G. Mitochondrial DNA sequence variation in Greeks. *Hum Biol* 2001; 73: 855-69.
- [15] Comas D, Calafel F, Mateu E, *et al.* Geographic variation in human mitochondrial DNA control region sequence: The population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 1996; 13: 1067-77.
- [16] Schneider S, Roessli D, Excoffier L: *Arlequin ver. 2.00: A Software for Population Genetic Data Analysis*. Switzerland: Genetics and Biometry Laboratory, University of Geneva; 2000.
- [17] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; 123: 585-95.
- [18] Fu XY. Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. *Genetics* 1997; 147: 915-25.
- [19] Rogers AR, Harpending HC. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 1992; 9: 552-69.
- [20] Bandelt HJ, Forster P, Sykes BC, *et al.* Mitochondrial portraits of human populations using median networks. *Genetics* 1995; 141: 743-53.
- [21] Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; 16: 37-48.

- [22] Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 1979; 76: 5269-73.
- [23] Rohlf FJ: *NTSYSpc Numerical Taxonomy System, ver. 2.1*. Setauket, NY: Exeter Publishing, Ltd; 2002.
- [24] Ota T: *DISPAN: A Software for Genetic Distances and Neighbor-Joining Calculations*. University Park, PA: Pennsylvania State University; 1993.
- [25] Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press; 1987.
- [26] Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* 1983; 19: 153-70.
- [27] Kingman JFC. The coalescent. *Stochastic Processes and their Applications* 1982; 13: 235-248.
- [28] Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 1991; 129: 555-62.
- [29] Ray N, Mathias C, Excoffier L. Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol* 2003; 20: 76-86.
- [30] Zlojutro M. *Mitochondrial DNA Variation in Yakutia: The Genetic Structure of an Expanding Population*. MA Thesis. Lawrence, KS: University of Kansas; 2005.
- [31] Richards M, Macaulay V, Hickey E, *et al*. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000; 67: 1251-76.
- [32] Helgason A, Nicholson G, Stefánsson K, *et al*. A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet* 2003; 67: 281-97.
- [33] Ward RH, Frazier BL, Dew-Jager K, *et al*. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 1991; 88: 8720-4.
- [34] Sherry ST, Rogers AR, Harpending H, *et al*. Mismatch distributions of mtDNA reveal recent human population expansions. *Human Biology* 1994; 66: 761-75.
- [35] Zlojutro M, Rubicz R, Devor EJ, *et al*. Genetic structure of the Aleuts and circumpolar populations based on mitochondrial DNA sequences: a synthesis. *Am J Phys Anthropol* 2006; 129: 446-64.
- [36] Bamshad M, Kivisild T, Watkins WS, *et al*. Genetic evidence on the origins of Indian caste populations. *Genome Res* 2001; 11: 994-1004.
- [37] Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 1964; 29: 1-27.
- [38] Zvelebil M: The rise of the nomads in Central Asia. In Sherratt A. ed *The Cambridge Encyclopedia of Archaeology*. New York: Crown; 1980.
- [39] Anthony DW, Brown DR. Birth of the chariot. *Archaeology* 1995; 28: 36-41.
- [40] Cavalli-Sforza LL, Menozzi P, Piazza A. *The History and Geography of Human Genes*. Princeton: Princeton University Press; 1994.
- [41] Ismagulov O. *Ethnic Anthropology of Kazakhstan*. Alma Ata: Academy of Science of Kazakhstan; 1982.
- [42] Bowles GT: *The Peoples of Asia*. London: Weidenfeld & Nicolson; 1977.
- [43] Richards MB, Macaulay VA, Bandelt HJ, *et al*. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 1998; 62: 241-60.
- [44] Forster P, Harding R, Torroni A, *et al*. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 1996; 59: 935-45.
- [45] Herodotus. *The Histories* translated by De Selincourt A. New York: Penguin Classics; 1996.
- [46] Ruhlen M. *A Guide to the World's Languages*. Stanford: Stanford University Press; 1991.
- [47] Mair VH. Prehistoric Caucasoid corpses of the Tarim Basin. *J Indoeur Stud* 1995; 23: 281-370.
- [48] Lalueza-Fox C, Sampietro ML, Gilbert MTP, *et al*. Unraveling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. *Proc R Soc Lond B* 2003; 271: 941-7.
- [49] Seielstad M, Minch E, Cavalli-Sforza LL. Genetic evidence for a higher female migration rate in humans. *Nat Genet* 1998; 20: 278-80.
- [50] Anderson S, Bankier AT, Barrell BG, *et al*. Sequence and organization of the human mitochondrial genome. *Nature* 1981; 290: 457-65.

Received: November 26, 2008

Revised: December 23, 2008

Accepted: January 20, 2009

© Devor *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.