

DNA Diversity and Population Admixture in Anatolia

Giulietta Di Benedetto,¹ Ayşe Ergüven,² Michele Stenico,¹ Loredana Castrì,³ Giorgio Bertorelle,² Inci Togan,² and Guido Barbujani^{1*}

¹*Dipartimento di Biologia, Università di Ferrara, I-44100 Ferrara, Italy*

²*Department of Biology, Middle East Technical University, 06532 Ankara, Turkey*

³*Dipartimento di Biologia Evoluzionistica e Sperimentale, Università di Bologna, Bologna, Italy*

KEY WORDS gene flow; mitochondrial DNA; Y chromosome; microsatellites; languages

ABSTRACT The Turkic language was introduced in Anatolia at the start of this millennium, by nomadic Turkmen groups from Central Asia. Whether that cultural transition also had significant population-genetics consequences is not fully understood. Three nuclear microsatellite loci, the hypervariable region I of the mitochondrial genome, six microsatellite loci of the Y chromosome, and one *Alu* insertion (*YAP*) were amplified and typed in 118 individuals from four populations of Anatolia. For each locus, the number of chromosomes considered varied between 51–200. Genetic variation was large within samples, and much less so between them. The contribution of Central Asian genes to the current Anatolian gene pool was quantified using three different methods, considering for comparison populations of Mediterra-

nean Europe, and Turkic-speaking populations of Central Asia. The most reliable estimates suggest roughly 30% Central Asian admixture for both mitochondrial and Y-chromosome loci. That (admittedly approximate) figure is compatible both with a substantial immigration accompanying the arrival of the Turkmen armies (which is not historically documented), and with continuous gene flow from Asia into Anatolia, at a rate of 1% for 40 generations. Because a military invasion is expected to more deeply affect the male gene pool, similar estimates of admixture for female- and male-transmitted traits are easier to reconcile with continuous migratory contacts between Anatolia and its Asian neighbors, perhaps facilitated by the disappearance of a linguistic barrier between them. *Am J Phys Anthropol* 115:144–156, 2001. © 2001 Wiley-Liss, Inc.

Evolutionary inferences from data on contemporary populations are complicated by the fact that long-term population sizes, rates of gene-flow, and selection coefficients are seldom known. As a consequence, observed patterns of genetic variation are often compatible with more than one evolutionary model. To discriminate among competing hypotheses, however, one can exploit the available wealth of archaeological and linguistic data. Comparative studies of biological and cultural diversity (Sokal, 1988; Cavalli-Sforza et al., 1988; Torroni et al., 1993; Barbujani and Pilastro, 1993; Ward et al., 1993; Sajantila et al., 1995; Poloni et al., 1997) have provided insights into important aspects of the human evolutionary history.

In Europe, many linguistic barriers, and especially those between Indo-European and non-Indo-European speakers, are associated with increased genetic change (Sokal et al., 1990, and references therein), probably because genetic and linguistic diversity have often been shaped by the same demographic changes (Cavalli-Sforza et al., 1988; Barbujani, 1997). However, despite the presence of a major language barrier (Altaic vs. Indo-European), no clear discontinuity has been described so far between the European and the Anatolian (i.e., Asian Turkish) gene pools, in large-scale analyses of blood groups, electrophoretic polymorphisms (Sokal et al., 1988; Harding and Sokal, 1988; Simoni et al., 1999),

and mitochondrial DNA (Comas et al., 1996; Simoni et al., 2000). Conversely, mitochondrial data suggest that a statistically significant difference exists between Anatolia and its Arabic-speaking southern neighbors (Simoni et al., 2000). The limited genetic change across the boundary between Turkic and Indo-European languages calls for an explanation. Has linguistic change occurred independently from genetic change in Anatolia, and, if so, why?

To address this question, we collected blood samples in four Anatolian population groups, from which 11 DNA markers were typed. By analyzing those data, along with other data collected in the literature, we then tested whether DNA variation in Anatolia is consistent with any of the demographic models that can be built on the basis of historical and linguistic evidence.

Grant sponsor: Italian Ministry of Universities; Funds: COFIN 1999–2001; Grant sponsor: University of Ferrara; Grant sponsor: Turkish Scientific and Technical Council.

*Correspondence to: Guido Barbujani, Dipartimento di Biologia, Università di Ferrara, via L. Borsari 46, I-44100 Ferrara, Italy. E-mail: bjj@unife.it

Received 29 June 2000; accepted 6 March 2001.

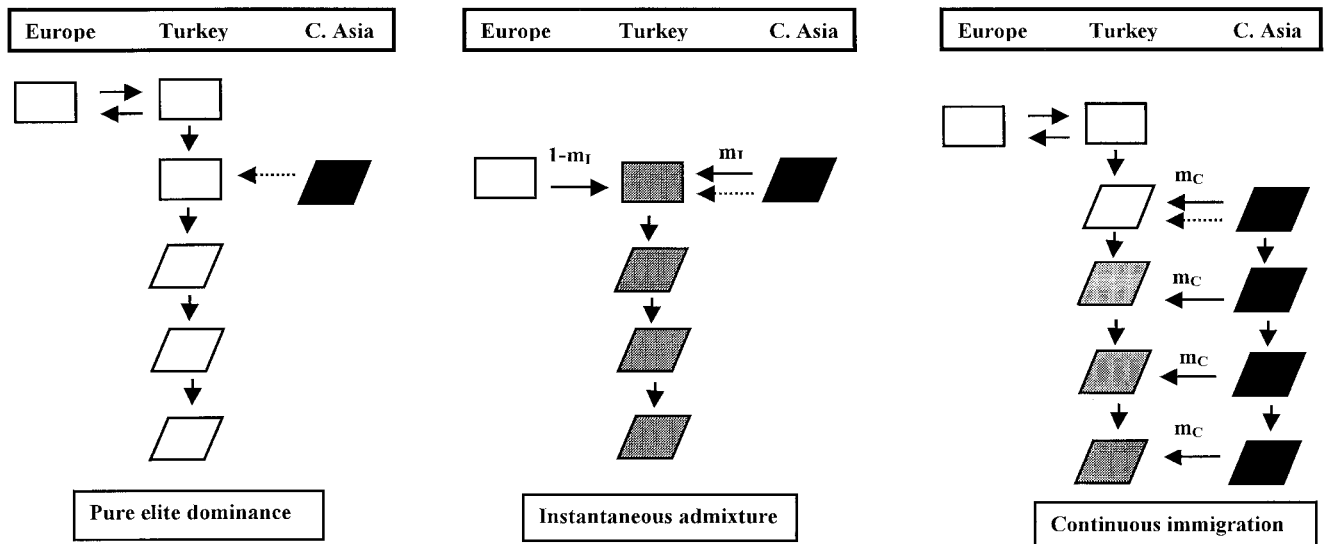


Fig. 1. Schematic representation of the three models tested against DNA data in this study. Rectangles are Indo-European-speaking populations; lozenges are Turkic-speaking populations. Dashed arrows represent linguistic transformations, horizontal solid arrows indicate gene flow, and vertical solid arrows indicate inheritance, from older (top) to younger (bottom) generations. Different shades of gray represent the likely proportion of alleles of Central Asian provenance in the Turkish allele pool.

THREE MODELS

The historical record shows that, in the 11th century AD, Anatolia was invaded by nomadic groups from Central Asia, collectively referred to as Oghuz (Akyildiz, 1997; Endress, 1988). The Oghuz Turks, called the Turkmen in Europe, are documented in the area between Mongolia and the Caspian sea from the 9th century AD. Under the leadership of the Seljuq family, they entered Iran, and, in 1044, emerged as secular rulers of the entire Islamic Near East, except Syria and Egypt. With their invasion of Anatolia in 1071 (Roux, 1984; Akyildiz, 1997; Endress, 1988), their language was imposed upon most resident populations, previously speaking Indo-European languages (Ruhlen, 1991).

The linguistic and political consequences of these episodes are well-documented (e.g., Renfrew, 1987), but it is not clear to what extent the Anatolian gene pool was affected by the Oghuz invasion. Schematically, three main scenarios may be envisaged, and they are liable to be tested using genetic data. One possibility is *elite dominance* (Renfrew, 1989), i.e., the process whereby the language of a few individuals is adopted by the rest of the population. Such elites are often military, and in their scrupulous study of the effects of historical episodes on population affinities, Sokal et al. (1996) concluded that military attacks have very limited genetic consequences. For the sake of clarity, here we shall posit that they had no genetic consequence at all (*pure elite dominance*; Fig. 1); indeed, there is evidence among Finn speakers of language replacement with no detectable genetic consequences (Sajantila and Pääbo, 1995). An alternative is that the arrival in Anatolia of substantial numbers of Central Asian Turkic-speakers would have caused parallel linguis-

tic and demographic changes, the latter reflected in new allele frequencies and in the presence of novel alleles in the Anatolian gene pool. That may have happened either at a specific moment in time (second possibility: *instantaneous admixture*) or through *continuous immigration* across many generations (and that is the third possibility). Because the Oghuz invaders were soldiers, and therefore mostly males, if immigration was instantaneous, greater effects may be expected upon Y-chromosome variation, whereas a continuous immigration seems compatible with equally large changes in the female- as well as in the male-transmitted portion of the genome. In this study we looked for evidence supporting any of the three models outlined above (Fig. 1), whose consequences are summarized in Table 1. Note that in all three cases, the linguistic transition may be explained by the imposition of the language of a minority to all members of the Anatolian population, i.e., by elite dominance sensu Renfrew (1989).

MATERIALS AND METHODS

Samples

Blood samples were collected from 118 male, Turkic-speaking, unrelated blood donors dwelling in four areas of Anatolia (Fig. 2), namely the Aegean coast around the city of Izmir (sample IZM), the southern Mediterranean coast around the town of Antalya (sample ANT), the central Anatolian plain around Ankara (sample ANK), and eastern Anatolia, close to the lake of Van (sample VAN). For ethical and legal reasons, the members of the samples were anonymous. Inhabitants of the major towns, who have a higher probability to be recent immigrants, were excluded, and special care was taken to avoid related individuals. The blood was preserved

TABLE 1. Summary of demographic models tested

Model	Brief explanation	Expected genetic consequences	
		Contribution of Asian alleles	Effects on Y-chromosome diversity
Pure elite dominance	Language change not associated with significant demographic change	Zero	None
Instantaneous admixture ¹	Language change due to demographic change	Greater than zero	Greater than on mtDNA diversity
Continuous immigration ¹	Language change followed by demographic change	Greater than zero	Same as on mtDNA diversity

¹ These models were termed, respectively, "intermixture" and "gene flow" by Long (1991).

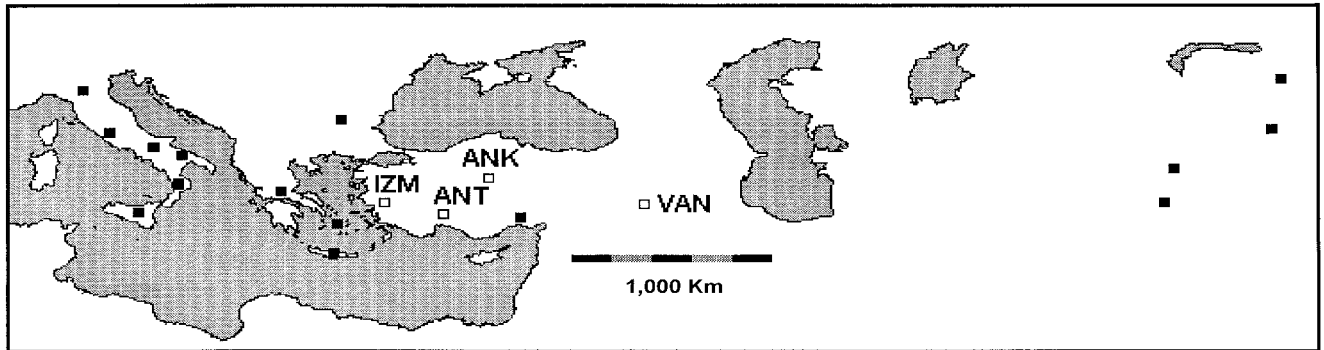


Fig. 2. Locations of the four Turkish samples (open squares), and of other samples used for estimating admixture (solid squares). IZM, Aegean coast; ANK, central Anatolia; VAN, eastern Anatolia; ANT, Mediterranean coast.

in K-EDTA solutions, and it was kept at -80°C until DNA extraction. No information was recorded about self-assessed ethnic affiliations.

DNA extraction and mitochondrial sequencing

DNA was extracted from the whole blood, either by the classical protocol (Maniatis et al., 1982) or by using a 5% Chelex-100 solution (Walsh et al., 1991). It was then suspended in sterile water and kept at 4°C . The mitochondrial control region was amplified using the primers L15926 and H408. A sequence of 360 base pairs within the first hypervariable region (hereafter, HVRI), between positions 16024–16383 of the Cambridge reference sequence (CRS; Anderson et al., 1981) was obtained by Thermo SequenaseTM cycle sequencing (Amersham kit), using L15996 and/or H16401 (Vigilant et al., 1989). Variable positions in the sequence are here indicated by the numbering of the CRS less 16,000. The sequences of the primers are in Table 2.

Nuclear microsatellites

Three microsatellite loci were analyzed. PLA2A (Hammond et al., 1994) and MFD179 (Deka et al., 1995) were amplified using steps of 94°C for 30 sec, annealing steps of 58°C for 30 sec, and elongation steps of 72°C for 30 sec. TH01 (Hammond et al., 1994) was amplified under similar conditions, but increasing the annealing temperature to 61°C . Thirty cycles were performed for all loci. The reactions were terminated by a step of 72°C for 3 min. (primers are listed in Table 1). The alleles (here designated by their repeat number, but see footnote

of Table 2) were separated on an 8% acrylamide gel, and the bands were visualized by a silver staining procedure.

Y-chromosome markers

Seven loci of the Y chromosome were typed, six of them microsatellites. An *Alu* insertion, or *YAP* element, was amplified in 35 cycles, using a denaturation step of 94°C for 30 sec, an annealing step of 51°C for 30 sec, and an elongation step of 72°C for 45 sec (Hammer and Horai, 1995). DXYS156 was amplified in 35 cycles, using a 58°C annealing temperature for 30 sec (Chen et al., 1994). Four tetranucleotide repeat loci, DYS19, DYS390, DYS391, and DYS393, were amplified together in 35 cycles, using denaturation steps of 94°C for 30 sec and elongation steps of 72°C for 1 min; a 58°C annealing step was performed for 20 cycles, and then the temperature was decreased to 56°C for the last 15 cycles (Kaiser et al., 1997). Finally, a trinucleotide microsatellite, DYS392, was independently amplified under the same conditions as above. All reactions were terminated by a step of 72°C for 5 min (sequences of primers are in Table 1).

The presence of the *Alu* insertion was investigated by running the amplified products in 2% agarose gel, and checking whether a 455-bp or a 150-bp band was present. The alleles of the microsatellite loci were visualized by an automated sequencing system (ALF-Pharmacia), using CY5 labelled primers, or silver staining dye after a run over a 6% acrylamide gel.

TABLE 2. Primers used for amplification

Locus	Primer name	Primer sequence	Reference
HVR-I	L15926	5'-tacaccagtcttgtaaacc-3'	Present study
	H408	5'-ctgttaaaagtgcataccgcca-3'	
	L15996	5'-ccaccattagcaccctaaagct-3'	
	H16401	5'tgattttcacggaggatgggtg-3'	
DYS-19(27h39)	Forward	5'-ctactgagtttctgttatagt-3'	Kaiser et al., 1997
	Reverse	5'-atggcatgtagtgaggaca-3'	
DYS-390	Forward	5'-tatattttacacatttttgggcc-3'	Kaiser et al., 1997
	Reverse	5'-tgacagtaaaatgaacacattgc-3'	
DYS-391	Forward	5'-ctattcattcaatcatcaccca-3'	Kaiser et al., 1997
	Reverse	5'-gattcttttggtgggtctg-3'	
DYS-392	Forward	5'-tcattaatctagcttttaaaaacaa-3'	Kaiser et al., 1997
	Reverse	5'-agaccagttgatgcaatgt-3'	
DYS-393	Forward	5'-gtggctcttctacttgtgtcaatac-3'	Kaiser et al., 1997
	Reverse	5'-aactcaagtccaaaaatgagg-3'	
DXYS156	Forward	5'-gtagtggctctttgcctcc-3'	Chen et al., 1994
	Reverse	5'-cagataccaaggtgagaatc-3'	
YAP	Forward	5'-caggggaagataaagaaata-3'	Hammer and Horai, 1995
	Reverse	5'-actgctaaaaggggatggat-3'	
TH01	Forward	5'-gtgggctgaaaagctcccattat-3'	Hammond et al., 1994
	Reverse	5'-attcaaaagggatctctggctctgg-3'	
PLA2A	Forward	5'-ggttgtaagctccatgaggtaga-3'	Hammond et al., 1994
	Reverse	5'-ttgagcacttactatgtgccaggct-3'	
MFD179	Forward	5'-caaattcaattcttccagc-3'	Deka et al., 1995
	Reverse	5'-actgtactcctgcatgtag-3'	

Diversity indices and phylogenetic analysis

The nucleotide diversity π was estimated as

$$\pi = \frac{n}{n-1} \frac{1}{L} \sum_j \left(1 - \sum_i x_{ij}^2 \right)$$

where n is the sample size, L is the length of the sequence of interest, and x_{ij} is the frequency of the i -th nucleotide at the j -th site (Nei and Tajima, 1981).

Gene diversity D (often referred to as heterozygosity) was estimated as

$$D = (n/n - 1)(1 - \sum q_i^2)$$

where n is sample size and q_i is the allele frequency (Nei, 1987). Genetic heterogeneity, within and among populations, was quantified by estimating Wright's F statistics, or their molecular equivalents, the Φ statistics, by means of the analysis of molecular variance (AMOVA; Excoffier et al., 1992). The Φ statistics are genetic variances, estimated taking into account not only frequency differences between individuals and populations, but also the molecular differences between mitochondrial sequences or microsatellite alleles. The significance of the variance components thus estimated was assessed by randomization (Excoffier et al., 1992).

The phylogenetic relationships among sequences were inferred by maximum parsimony. One thousand putatively most-parsimonious trees were generated in repeated runs of DNAPARS* from Phylip 3.5c (Felsenstein, 1989), and the consensus tree was obtained by CONSENSE from the same software package.

Estimating admixture: assumptions

Models of admixture are necessarily based on several assumptions (e.g., see Guglielmino et al., 1990; Long, 1991; Comas et al., 1998), which we shall here make explicit. To quantify the contribution of alleles of Central Asian provenance to the current Anatolian gene pool, the simplest starting point is to assume that, before the 11th century AD, the Turkish population was genetically similar to its neighbors in the Mediterranean area, i.e., the Levant and Southeastern Europe. If so, the best estimates of Turkish allele frequencies before admixture are the frequencies of the same alleles in the available samples from that area, q_E .

Second, before 992 AD, the Oghuz tribes were located in what is now Kirghizistan (Endress, 1988). Today, the Kazakh, Uighurs, and Kirghiz of Central Asia are among the closest linguistic relatives of the Turks, all belonging to the Common Turkic branch of the Altaic language phylum (Ruhlen, 1991). If linguistic affinities reflect common origins in the not-so-remote past (as suggested by several authors, including Sokal, 1998; Cavalli-Sforza et al., 1988, 1994), the best available estimate of the immigrants' allele frequencies are the frequencies of the same alleles among Turkic speakers of Central Asia, q_A .

The genetic data available for Kirghiz, Kazakh, and Uighur (Comas et al., 1998; Perez-Lezaun et al., 1999) seem therefore suitable for this purpose. We believe that this is true even if these populations show some evidence of admixture, which Comas et al. (1998) attributed to their central position in a gradient going from East Asia to Europe, and/or to the movement of people along the Silk Road (Comas et al., 1998). In both cases, the gene-flow phenomena

that could have affected the genetic composition of the Kirghiz, Kazakh, and Uighur appear older and more time-diluted than the specific admixture process we are analyzing in this study. Given the allele frequencies estimated in this study for the four Turkish samples, q_T , the contribution of Central Asian genes to the Anatolian gene pool was thus estimated in three ways.

The first two methods are classical ways to quantify admixture in populations where migration can be considered unidirectional (from Asia into Turkey, in this case). Both methods are applied to allele frequencies, and the results are averaged across alleles. Since we also had measures of *sequence* divergence between alleles, we also used a third, multilocus method for inferring admixture from the mean coalescence times between alleles.

Estimating admixture: 1. Single-locus, instantaneous gene flow. Under an island model (Wright, 1969), and assuming that all gene flow from Central Asia into Turkey occurred at one moment in time, the relationship between the allele frequencies in Turkey, Europe, and Central Asia is

$$q_T = (1 - m_I)q_E + m_Iq_A$$

where m_I is the only unknown term, and can be defined as the instantaneous gene-flow rate, i.e., the amount of immigration necessary for q_E to reach q_T in one generation's time.

Estimating admixture: 2. Single-locus, continuous gene flow. Alternatively, one may imagine that the current allele frequencies in Anatolia are the result of continuous gene flow from Asia, across the 1,000 years, or 40 generations, following the Oghuz invasion. From the formula above, one obtains

$$q_T = q_A + (1 - m_C)^{40} \times (q_E - q_A)$$

which can be solved for m_C , the rate of continuous gene flow per generation. Note that 40 generations is a minimum estimate of the time through which gene flow may have occurred between Anatolia and Central Asia, once the language barrier between them began to be removed. Had genetic exchange begun earlier, lower rates of gene flow (i.e., lower m_C values) could account for the data.

Estimating admixture: 3. Multilocus, instantaneous gene flow. To estimate admixture from multilocus data (see also Chakraborty et al., 1992), we chose a recently developed method based on the estimation of the mean coalescence times between pairs of alleles drawn from the three populations of interest (Bertorelle and Excoffier, 1998). The molecular differences between alleles are considered under an infinite-site model or, for microsatellite data, under a single-step stepwise mutation model. The relative weight of alleles coming from Central Asia in the composi-

tion of the hybrid population's gene pool, m_M , was estimated from those coalescence times. (In Bertorelle and Excoffier (1998), m_M is called m_Y , a symbol we preferred to avoid here because of the ambiguity with the Y-chromosome data; in the same paper, m_c is another estimator of admixture, which was not used in the present study.)

Other samples considered and expected results

Allele frequencies of TH01 in an additional Anatolian sample, Adana (Alper et al., 1995), were added to the database analyzed, for a total of 590 chromosomes. The same was done with 74 Turkish mitochondrial sequences coming from all over the country (Comas et al., 1996; Calafell et al., 1996), bringing the total Turkish mtDNA sample size to 146. Data about Central Asian populations are from Comas et al. (1998) for mtDNA, and from Perez-Lezaun et al. (1999) for Y-chromosome polymorphisms. In addition, we had unpublished allele frequencies for TH01 in 11 Uighurs and 9 Kirghiz (Luiselli et al., personal communication). The q_E values of this study, for both nuclear and mitochondrial loci, are based on samples from Bulgaria, Greece, Crete, peninsular Italy, and Sicily (Fig. 2). To the best of our knowledge, samples from the Levant have only been typed for different Y-chromosome markers than those typed in this study (Scozzari et al., 1997; Semino et al., 2000). Therefore, for the sake of consistency, although mtDNA data were available in samples from the Levant (Druzes and Near-Eastern Arabs), they were not considered here.

Clusters of mitochondrial alleles

A mitochondrial phylogeny for Eurasia as a whole is not established yet, and it is unclear which sites are most informative for identifying evolutionary relationships among sequences from the two continents. Based on 217 individual sequences, Comas et al. (1998) listed nine substitutions which appeared restricted to Asia and six which were only found in Europe. That classification proved unsuitable in the present study, for some European sequences showed substitutions defined as typically Asian (e.g., at 16189), and vice versa (e.g., the 16129–16223 motif). We therefore decided to identify in the data clusters of mtDNA sequences that show likely phylogenetic relationships, and to use their frequencies for comparing the three population groups of interest. Most such clusters correspond to the haplogroups which have been separately identified for Europe (Macauley et al., 1999) and for Asia (Torroni et al., 1993; Kolman et al., 1996; Starikovskaya et al., 1998; Castri, unpublished data).

RESULTS

In the first hypervariable region of mtDNA, a 360-bp sequence was typed in 72 individuals (17, 19, 16, and 20 for the ANK, ANT, IZM, and VAN sam-

ples, respectively). Seventy-four polymorphic sites defined 61 haplotypes (Fig. 3), with a gene diversity $D = 0.957$. D is approximately constant for the four samples (0.923 in ANK, 0.886 in ANT, 0.933 in IZM, and 0.942 in VAN), and the mean number of pairwise differences is also almost identical in ANK, ANT, and VAN (respectively, 4.69, 4.74, and 4.65), with a lower value (3.47) only in IZM. As a consequence, estimates of nucleotide diversity, π , are 0.013 in ANK, ANT, and VAN, and 0.010 in IZM. The most frequent sequence in Anatolia is the CRS (which is also the most common sequence in Europe) with a total frequency of 0.139. No significant differentiation is evident among samples ($\Phi_{ST} = 0.004$, n.s.), and neither was that variance significant when two other samples (Calafell et al., 1996; Comas et al., 1996) were compared.

Anatolian mtDNAs were classified in 24 groups based on HVR I motifs (Table 3). Some European and Asian haplogroups defined in previous studies (such as haplogroups H and G, and some subgroups of cluster UK) were ignored due to the impossibility of identifying them on the basis of HVRI sequences. Even then, some ambiguities remained, because some clusters are characterized by substitutions at sites that may have undergone recurrent mutations (e.g., 16189, 16304, 16362, or 16311; Macalauy et al., 1999). This is the case of haplogroup R1, which is defined on HVRI by a rather widespread transition at position 16311, and cluster JT, which is characterized by a transition at 16126. Moreover, some mutations have been observed in association with various haplogroups, whose frequency may consequently be under- or overestimated (e.g., cluster M). An example is the Asian haplogroup E, which shares with the European haplogroup X two transitions (16223 and 16278), also found in African sequences (Rando et al., 1998).

Despite these ambiguities, the Anatolian populations tend to show intermediate frequencies between South European and Central Asian values (Fig. 4). Note the relatively high frequency of sequences belonging to J, a rather common haplogroup in the Near East, and the presence of what Richards et al. (1998) termed subcluster J1b1, i.e., the HVRI motif 16069-16261-16145-16222-16172 (Fig. 3), so far observed mostly in Northern Europe, but also in Italy (Richards et al., 1998) and Central Asia (Comas et al., 1999).

Maximum-parsimony trees were estimated to assess the reliability of clusters of mitochondrial alleles used in this study. The resulting consensus tree (Fig. 5) confirmed the subdivision into the 24 groups listed in Table 3. Two main clusters are apparent in that tree. The first one only includes two sequences likely related to the L3a* group (HVRI motif 16145-16176G-16223), observed in Africa, and which has candidate members in Southern Europe and the Near East (Macalauy et al., 1999). Most other sequences fall in a much larger cluster, within which several previously de-

scribed haplogroups can be identified. Another evident split separates the alleles harboring the 16223T substitution (Asian haplogroups M, C, and D, and European haplogroups I and X) from the rest of the sample.

We could attribute to the 24 haplogroups (Table 3) 62.3% of the Turkish mtDNA sequences available, 69.3% of the 205 Central Asian sequences, and 63.4% of the 142 European sequences. For the two single-locus methods described above, admixture proportions were inferred from the frequencies of these haplogroups.

Y-chromosome markers were studied in 118 individuals (Table 4), but the complete set of loci could only be typed in 51 of them, among which 45 different multilocus haplotypes were found. D is lowest in the ANK sample (0.40), whereas in ANT, IZM, and VAN, D was respectively equal to 0.54, 0.53, and 0.47. The ANK sample shares one haplotype with IZM and another with VAN. No other shared haplotypes were observed. Once again, molecular variances are not significantly greater than zero among samples ($\Phi_{ST} = 0.027$, n.s.).

Two hundred chromosomes, 50 for each sample, were typed for the autosomic loci (Table 4). Contrary to what observed for mtDNA and the Y chromosome, AMOVA shows significant differentiation among samples, accounting for only 1.82% of the total genetic variance, but reaching the $P = 0.011$ significance level ($\Phi_{ST} = 0.018$). Because genetic variances were tested independently three times (for mitochondrial, Y-chromosome, and autosomic loci), this level of significance must be multiplied by 3. After such a Bonferroni correction (Sokal and Rohlf, 1995), the overall differentiation remains significant ($P = 0.033$), but that significance is entirely due to the effect of the MFD179 locus, where diversity among populations accounts for 7.4% of the total, $\Phi_{ST} = 0.070$ ($P < 0.001$). On the contrary, among-population variances are insignificant for TH01 ($\Phi_{ST} = 0.013$, $P = 0.069$) and for PLA2A ($\Phi_{ST} = 0.008$, $P = 0.164$).

Preliminary treatment of data for inferring admixture

The relative contribution of Central Asian genes to the current Turkish gene pool could be estimated on the basis of the mitochondrial HVRI, TH01, DYS390, DYS391, DYS392, DYS393, and DYS19. European and/or Central Asian data about MFD179, PLA2A, DYXS156, and YAP were insufficient for the purposes of the analysis, and had to be excluded. The Turkish samples described by Comas et al. (1996) and Calafell et al. (1996) were also insignificantly different from ours, as well as the TH01 frequencies of the Turkish population of Adana (Alper et al., 1996). Therefore, in all successive steps of the analysis we treated all Turkish data as a single entity, thus assuming that the heterogeneity observed at the

1111111111	1111111111	1111111111	1111111111	1111111111	1111111111	1111111111	1111111111	1111	I	A	A	V
6666666666	6666666666	6666666666	6666666666	6666666666	6666666666	6666666666	6666666666	6666				
0000011111	1111111111	1222222222	2222222222	2222222222	2222222222	2222222222	2222222222	3333	Z	N	N	A
6668902224	4467778899	9001122223	4444455566	6777788899	9999900000	1122222345	5556		M	T	K	N
6796352690	5881263912	3593823494	4568906814	5034878901	2456801349	1924578831	4782					
CRS	ACCTTTATGT	GCCATCATCC	CCTGCCCTTC	GCACTCCACC	ACGGCCTACC	CCCCTACGTA	TGATTCACAA	CTCT	1	5	2	2
	A...G...	..C...T...							0	0	1	0
	A...G...T...		T				0	0	0	1
	T.....		C.....T.....	..CT.		0	0	1	0
	..C.....T.....				C		0	0	2	0
A.....T.....T			CA.....	..C		0	1	0	0
G.....T.....			C.....	..C		1	0	0	0
	T.....T.....						0	0	0	1
	T.....		T	A.....	..A.....		1	0	0	0
	..C...A.T.....				T.....			0	0	0	1
G.A.C...T.....				T.....	C.....		0	1	0	0
	CC..T.....T.....					0	1	0	0
	CC..T.....T.....					0	1	0	0
	T...CC..T.....T.....					1	0	0	0
	C.....T.....T.....		C.....		0	0	0	1
	C.....T.....T.....		G.....		0	1	0	0
C.....C.....							1	0	0	0
C.....				TT.....				1	0	0	0
	..T...C..								0	0	0	1
	..T...C..	T..A.....T.....		G.....			0	1	0	0
	..T...C..					T		1	0	0	0
	..T...C..	A...C.....T.....T.....					0	0	0	1
	..T...C..	A.....TA.....T.....T.....					0	0	0	1
	..T...C..	A.....TA.....T.....T.....		C.....		0	0	0	1
	..T..A.C..	A.....T.....T.....	G.....					0	0	1	0
	C.....					C.....		0	0	0	1
	C.....					C.....		0	1	0	0
	..T.C.....C.....T.....				C.....		0	1	0	0
	..T.C.....C.....T.....				C...CT.....		0	1	0	0
	..T.....C.....					C.....		0	1	0	0
	..T.....T.....T.....T.....					0	0	0	1
	..T.....								0	0	1	0
	..T.....				A.....				0	0	0	1
	T.....							1	0	0	0
	T.....T.....						1	0	0	0
A.....					C		0	0	1	0
A.....								0	0	1	0
	T.....T.....T.....		C.....		1	0	0	0
	T.....T.....T.....					0	1	0	0
	T.....T.....T.....					1	0	0	0
A.....C.....T.....			T.....C.....	1	0	0	0
	T.....C.....			T.....		0	1	0	0
	C.....				C.....		0	0	1	0
	C.....T.....			C.....		1	0	0	0
	C.....T.....			A.....		1	0	0	0
A.....T.....	A...T.....			A.....		0	0	0	1
	T.....			CC.....T.....			0	0	0	1
	C.....							0	0	0	1
	..T.....								0	0	1	0
	..T.....	A.....							0	0	1	0
	T.....							0	0	1	0
	C...T.....T.....T...T...T...			T.....		0	0	0	1
	C.....							1	0	1	0
	C.....				C.....		0	0	0	1
	C.....				C.....		0	1	0	0
	C.....T.....						0	1	0	0
T...C..T...C..CT.....C.....T.....				0	0	1	0
T...C..T...C..C.....						1	0	0	0
T...CC..								0	0	1	0
	G.....A.....CC..T.....						0	0	0	1

Fig. 3. HVRI sequences in four Turkish samples. Absolute frequencies of each sequence in the four localities are shown in the right columns.

TABLE 3. List of HVR I motifs, along with respective haplogroup attribution

	HVRI motif	Haplo- group
Central Asia	16223 C → T	M
	16223 C → T 16298 T → C 16327 C → T	C
	16223 C → T 16362 T → C	D
	16223 C → T 16227 A → G 16278 C → T	E
	16223 C → T 16290 C → T 16319 G → A	A
	16362 T → C	
	16189 T → C 16217 T → C	B
	16140 T → C 16189 T → C	B5
	16304 T → C	F
Europe		CRS
	16298 T → C	V
	16067 C → T	HV
	16189 T → C 16249 T → C	U1
	16343 A → G	U3
	16192 C → T 16256 C → T 16270 C → T	U5
	16093 T → C 16224 T → C 16311 T → C	K
	16126 T → C	JT
	16069 C → T 16126 C → T	J
	16126 T → C 16294 C → T 16296 C → T	T
	16126 T → C 16163 A → G 16186 C → T	T1
	16189 T → C 16294 C → T	
	16223 C → T 16292 C → T	W
	16189 T → C 16223 C → T 16278 C → T	X
	16129 G → A 16223 C → T	I
	16311 T → C	R1
	16145 G → A 16176 C → G 16223 C → T	L3a*

MFD179 locus does not indicate substantial population subdivision.

Estimates of admixture

All admixture models assume that the hybrid population (Anatolia in our case) has intermediate genetic characteristics between those of the parental populations. In practice, because of genetic drift and sampling effects, some allele frequencies of the supposedly hybrid population may be higher or lower than those of both parental populations, leading to single-allele estimates higher than 1 or lower than 0. We refer to those results as *implausible*. When estimating a rate of continuous gene flow, the logarithms of negative numbers are of course impossible to calculate; we refer to those results as *intractable*. In what follows, only the intractable results were disregarded, whereas means, medians, and standard errors were estimated from both plausible and implausible results, assigning equal weight to each locus.

Table 5 summarizes the admixture estimates obtained by the single-locus approaches. The average m_I and m_c values estimated from the frequencies of the 24 mitochondrial haplogroups differ sharply from the respective median values, revealing a strong effect of some statistical outliers. The calculations were repeated several times, each time excluding one or more haplogroups, until it became evident that some rare haplogroups exerted a disproportionate effect on the final result. The haplogroup we refer to as V, for instance, is present in few individuals of Southern Europe and absent elsewhere, yielding m_I and m_c esti-

mates = 0. We eventually chose to reestimate m_I and m_c based on four haplogroups whose frequency is polymorphic (i.e., higher than 0.05) in at least two of the samples considered, i.e., D, K, T, and the CRS (and excluding M and X, whose frequencies are probably overestimated). These estimates, more robust than the previous ones, indicate that the effects of an instantaneous immigration of 29.5% of Central Asian alleles can be obtained under continuous immigration at a rate $m_c = 0.01$ per 40 generation.

When all tractable information is considered (Table 5, five loci), the Central Asian contribution to the Anatolian gene pool appears lower for the Y chromosome than for mtDNA. By analogy to what had been done for mtDNA, we selected six alleles complying with the criterion of being polymorphic in at least two populations (DYS19*15, DYS19*17, DYS390*25, DYS392*11, DYS392*12, and DYS393*13). In this way, the standard errors are reduced, and the mean and median estimates differ only slightly. The m_I inferred from these alleles is slightly higher than that inferred from mtDNA data; if it reflects continuous immigration from Central Asia, the rate is $m_c = 0.01$ per generation.

Estimates of admixture from the only nuclear marker for which comparison is possible suffer from the limited size of the Asian sample, 40 chromosomes. One allele proved to be intractable, and all alleles longer than TH01*10 have not been observed in Turkey and in the small Central Asian sample, leading to m_I estimates = 1, despite the fact that they are also extremely rare in Europe ($q_E < 0.003$). When these alleles are neglected, the autosomal estimate of m_I is a low 0.078, which does not significantly differ from the mitochondrial and Y-chromosome estimates, because of the small sample size and of the associated large standard error. Intractable values result in a negative m_c estimate, suggesting a similarity between Turkish and European allele frequencies.

The multilocus approach based on coalescence times gives a range of m_M values, also listed in Table 5. The analysis of selected sets of alleles has a lesser impact on the m_M than on the m_I and m_c estimates for mitochondrial data, but not for Y-chromosome polymorphisms. Regardless of the number of haplogroups considered, female admixture seems close to 30%, in agreement with single-locus estimates. For Y-chromosome microsatellites, conversely, selection of alleles leads to a substantial decrease in the estimated Central Asian admixture. All the m_I and m_M estimates we consider reliable fall in the interval between 0.25–0.35, and all differ from zero by more than 2 standard errors.

In theory, by the multilocus method, one could also estimate the time elapsed since admixture. However, owing to the presence of the same alleles in the parental and in the hybrid populations, this estimate is zero both for mitochondrial and Y-chromosome polymorphisms (Bertorelle and Excoffier,

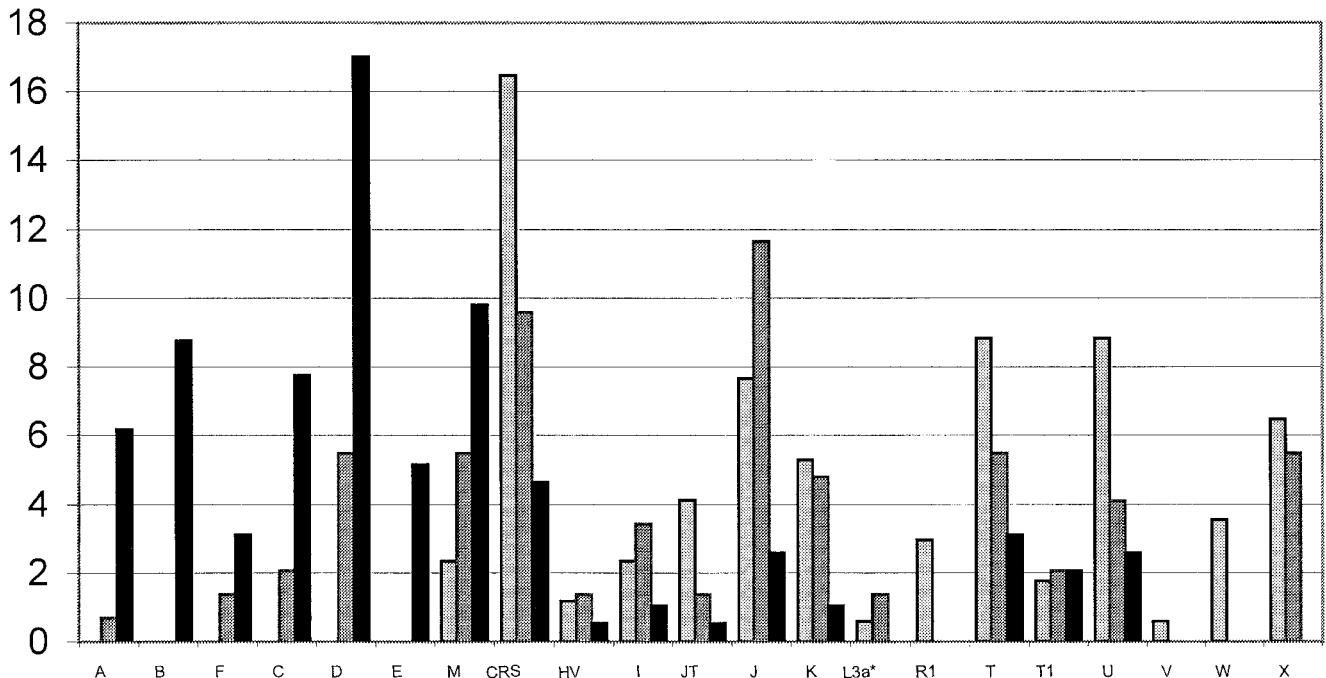


Fig. 4. Comparison of frequencies of various mitochondrial allele clusters or haplogroups in Mediterranean Europe (left, light grey bars), Anatolia (central, dark grey bars), and Turkic-speaking samples of Central Asia (right, black bars). Y-axis: percent values.

1998). Here we are speaking of units of mutational time, and so zero really means that admixture was recent enough not to be followed by successive differentiation, which is consistent with the limited time-scale, in evolutionary terms, of the events we are studying. This result is not to be taken at face value, but it certainly suggests that we are not dealing with the effects of very remote admixture.

DISCUSSION

As is usual with DNA data (Barbujani et al., 1997; Jorde et al., 2000), variation in Anatolia appears to be extensive within populations, and limited between them, with only one locus showing significant population differences. Contrary to what was observed elsewhere in Eurasia (e.g., Salem et al., 1996; Seielstad et al., 1998), population differences in Anatolia are not much greater for the Y chromosome than for mtDNA. As an example, genetic variances (F_{st}) inferred from Y-chromosome data in Central Asia were 40-fold higher than those inferred from mtDNA data (Comas et al., 1998; Perez-Lezaun et al., 1999), which was explained by an increased female mobility (Perez-Lezaun et al., 1999). In Anatolia things are clearly different. Estimates of Φ_{ST} among the four samples of this study are 5-fold higher for the markers of the Y chromosome than for the mtDNA markers, neither value being significantly greater than 0.

The models of admixture we know of do not incorporate the effects of genetic drift after admixture. Because drift and sampling have a heavier impact upon rare alleles, the estimates obtained from the common alleles are more robust; incidentally, only

for one such allele, DYS19*15, did the Anatolian allele frequencies fall out of the interval between their European and Central Asian counterparts (a result that we earlier defined as implausible).

The genetic features of populations before admixture are unknown, and must be approximated using information on contemporary samples (see Guglielmino et al., 1990). If the European populations of the eastern Mediterranean region are not too different genetically from the 11th century Anatolian population, and if the Turkmen incomers were not too different from the modern Turkic-speaking groups of Central Asia, this study shows that: 1) the Anatolian gene pool contains a substantial fraction of alleles of Asian origin; 2) immigration rates inferred from female- and male-transmitted traits are similar; 3) if there was a single, nearly instantaneous admixture event, some 30% of the current Anatolian genes have a Central Asian origin; and 4) if there was a continuous input of Central Asian alleles, it occurred at a rate of 1% per generation (or less, had the process started before the first Turkmen contact).

Admixture estimates have large standard errors, and it comes as no surprise that none of them differs significantly from the others. But it is interesting to note that, with one exception, the estimated m values converge in suggesting a Central Asian contribution to the current Turkish gene pool of around 30%. The exception is $m_M = 47\%$, based on Y-chromosome diversity. The m_M estimator is known to be biased towards 50% if the parent populations separated recently (Bertorelle and Excoffier, 1998). Because all other estimates we obtained are rather

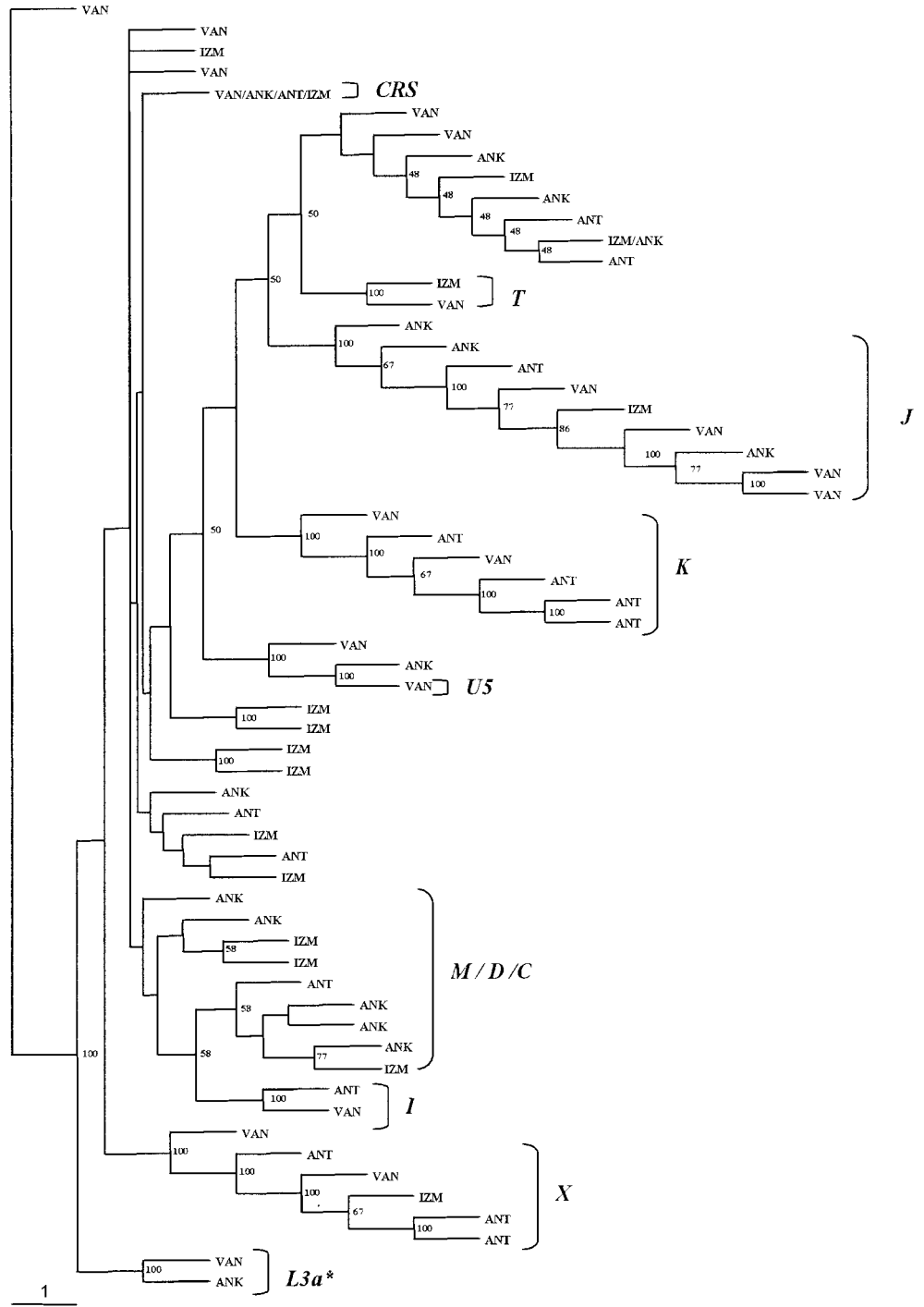


Fig. 5. Consensus maximum-parsimony tree of the Turkish mitochondrial sequences of this study. For each sequence, the geographic origin and, when possible, the haplogroup definition (according to Macaulay et al., 1999; Torroni et al., 1993; Kolman et al., 1996; Starikovskaya et al., 1998) are given. Figures at nodes refer to percentage of support for each branch.

consistent, we think it is safer to interpret the m_M estimate as due to the limited statistical power of our data for genealogical inferences over such a short time period.

An instantaneous input of Asian alleles, accounting for 30% of the current gene pool, means that the 11th century invasion entailed a massive movement of people, females as well as males. This is in contrast with historical reconstructions, referring to the Oghuz as an army or a tribe, and not as a large immigrating wave (Roux, 1984; Endress, 1988). Ge-

netic data cannot tell us whether the historical sources are reliable. But if most Asian alleles in the current Anatolian gene pool arrived in the 11th century AD, the Oghuz invasion had a much greater demographic impact than is commonly believed by historians.

The alternative is a continuous input of alleles from Central Asia (for the sake of clarity, it seems necessary to maintain the schematic opposition between instantaneous and continuous gene flow, although things may well have occurred in an inter-

TABLE 4. Allele frequencies of nuclear microsatellites in four Turkish samples¹

Locus	Allele	ANK	ANT	IZM	VAN
YAP	-	0.962	0.733	0.750	0.704
	+	0.038	0.267	0.250	0.296
	(N)	26	15	24	29
DYS156	11	0.000	0.062	0.130	0.053
	12	1.000	0.938	0.870	0.894
	13	0.000	0.000	0.000	0.053
	(N)	23	16	23	18
DYS19	11	0.036	0.000	0.000	0.034
	13	0.000	0.063	0.045	0.000
	14	0.000	0.063	0.227	0.103
	15	0.571	0.438	0.273	0.172
	16	0.250	0.188	0.409	0.552
	17	0.143	0.250	0.045	0.138
		(N)	28	16	22
DYS390	22	0.000	0.000	0.050	0.000
	23	0.050	0.188	0.100	0.000
	24	0.650	0.188	0.250	0.864
	25	0.200	0.563	0.600	0.136
	26	0.050	0.063	0.000	0.000
	27	0.050	0.000	0.000	0.000
		(N)	20	16	20
DYS391	9	0.105	0.000	0.000	0.133
	10	0.632	0.563	0.389	0.467
	11	0.263	0.438	0.611	0.400
	(N)	16	16	18	16
DYS392	10	0.056	0.000	0.056	0.000
	11	0.778	0.625	0.500	0.737
	12	0.111	0.125	0.056	0.053
	13	0.056	0.188	0.389	0.158
	14	0.000	0.000	0.000	0.053
	15	0.000	0.063	0.000	0.000
		(N)	18	16	18
DYS393	12	0.036	0.067	0.000	0.000
	13	0.571	0.533	0.636	0.607
	14	0.321	0.400	0.364	0.321
	15	0.071	0.000	0.000	0.071
		(N)	28	15	22
PLA2A	1	0.020	0.020	0.000	0.020
	2	0.520	0.580	0.420	0.440
	3	0.040	0.080	0.160	0.120
	4	0.180	0.100	0.120	0.040
	5	0.140	0.180	0.080	0.140
	6	0.040	0.020	0.200	0.140
	7	0.060	0.020	0.020	0.060
	8	0.000	0.000	0.000	0.040
	(N)	50	50	50	50
MFD179	2	0.000	0.000	0.000	0.040
	3	0.000	0.060	0.000	0.080
	4	0.000	0.020	0.040	0.100
	5	0.560	0.740	0.680	0.340
	6	0.240	0.020	0.220	0.320
	7	0.180	0.120	0.040	0.080
	8	0.020	0.000	0.020	0.020
	9	0.000	0.040	0.000	0.020
		(N)	50	50	50
TH01	6	0.120	0.220	0.260	0.240
	7	0.380	0.140	0.180	0.320
	8	0.200	0.140	0.160	0.080
	9	0.240	0.400	0.140	0.160
	9.3	0.000	0.000	0.100	0.040
	10	0.060	0.100	0.160	0.160
	(N)	50	50	50	50

¹ Microsatellite alleles labelled according to number of repeats they contain. For PLA2A and MFD179, allele labels are conventional figures (Hammond et al., 1994; Deka et al., 1995), from shortest to longest. (N), sample size.

mediate manner). Is it realistic to imagine 40 generations of gene flow from Central Asia into Anatolia, at a rate of $m = 0.01$?

In a comparable study, gene flow from non-Jewish neighbors into the Jewish gene pool was estimated between 0.6–2.3% per generation, for the last 100 generations (Morton et al., 1982). Higher figures, up to 8.7%, have been estimated among Italian communities of the Po valley (Barrai et al., 1984). Therefore, the immigration rate obtained for Anatolia is not unreasonably high for western Eurasia. However, the above populations moved within a rather small area. Although Asia and Europe were connected through Anatolia by one of the major medieval trading routes, the Silk road (see Comas et al., 1998), it is unclear whether Central Asian groups could consistently contribute as much as 1% of the Anatolian gene pool at each generation.

One possibility is that, once a Turkic language came to be spoken in Anatolia, gene flow from linguistically related areas was facilitated. Language barriers have been shown to reduce levels of gene flow in various regions of the world (see Sokal et al., 1990), including the Caucasus, at the borders of Turkey (Barbujani et al., 1994). However, the opposite case, i.e., higher genetic exchange between geographically distant but linguistically related groups, has only been observed in Africa (Excoffier et al., 1991). If the amount of admixture estimated in this study is due to continuous gene flow, long-range migration between linguistic relatives would appear substantial in this part of Asia as well.

Which of the models outlined in Table 1 seems to best account for the origin of the current Anatolian gene pool? The hypothesis that we called pure elite-dominance is contradicted by the fact that the Central Asian contribution to the Anatolian gene pool appears substantial, regardless of the numerical method used to quantify admixture. It seems worthwhile to emphasize that this result does not rule out that the linguistic replacement, in itself, was an episode of elite dominance, as defined by Renfrew (1989). What this study shows is that the Asian contribution to the Anatolian gene pool is not zero. Accordingly, two other possibilities remain. One is that the arrival of the Oghuz armies was more a large-scale population movement than a military invasion, contrary to what is suggested by the historical record. This is the model that we called instantaneous immigration. That model, however, predicts greater effects at the Y-chromosome than at the mtDNA level, which this study does not confirm. Alternatively, the historical record may be accurate in suggesting that small numbers of Oghuz Turks invaded Anatolia. In that case, continuous gene flow from Asia should be envisaged at a rate of around 1% per generation, i.e., what we termed a model of continuous immigration. A gene flow rate at around 1% for 40 generations represents a substantial migration process across the large distances separating Anatolia from Central Asia. Genetic exchange, however, may have been enhanced by linguistic relatedness, which may have weakened cultural barriers to immigration. At this stage, continuous im-

TABLE 5. Estimates of Central Asian admixture in Turkish gene pool¹

Polymorphism	Average $m_I \pm SE$	Median m_I	Average $m_C \pm SE$	Median m_C	$m_M \pm SE$
mtDNA, 24 haplogroups	0.210 \pm 0.080	0.033	0.008 \pm 0.004	0.001	0.301 \pm 0.105
mtDNA, 4 haplogroups	0.295 \pm 0.110	0.285	0.010 \pm 0.004	0.008	0.336 \pm 0.126
Y chromosome, 5 loci	0.157 \pm 0.200	0.147	0.006 \pm 0.006	0.005	0.469 \pm 0.105
Y chromosome, 6 alleles	0.319 \pm 0.043	0.314	0.010 \pm 0.001	0.009	0.259 \pm 0.091
TH01, 9 alleles	0.453 \pm 0.250	0.689	0.002 \pm 0.017	0.000	0.346 \pm 0.080
TH01, 5 alleles	0.078 \pm 0.077	0.020	-0.005 \pm 0.004	-0.004	

¹ Standard errors estimated assuming that allele frequencies are independent. m_I , single locus, instantaneous, based on allele frequencies; m_C , single-locus, continuous, based on allele frequencies; m_M , multilocus, instantaneous, based on a coalescent approach.

migration from Central Asia seems the model which is simplest to reconcile with the available data.

ACKNOWLEDGMENTS

We thank Lucia Simoni for many comments and for critical reading of this manuscript; Jaume Bertranpetit, Anna Perez-Lezaun, and Donata Luiselli for giving us access to unpublished material; and Peter De Knijff for sending us allele ladders. Preliminary work was carried out in Loredana Nigro's laboratory, at the University of Padua. Loredana passed away in October 1998; we miss her, and this paper is dedicated to her.

LITERATURE CITED

- Akyildiz E. 1997. Tas Çagından Osmanlı'ya Anadolu. Istanbul: A.D. Yayincilik.
- Alper B, Wiegand P, Brinkmann B. 1995. Frequency profiles of 3 STRs in a Turkish population. *Int J Legal Med* 108:110–112.
- Anderson S, Bankier T, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger S, Schreier PH, Smith AJH, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.
- Barbujani G, Pilastro A. 1993. Genetic evidence on origin and dispersal of human populations speaking languages of the Nostatic macrofamily. *Proc Natl Acad Sci USA* 90:4670–4673.
- Barbujani G, Nasidze IE, Whitehead GN. 1994. Genetic diversity in the Caucasus. *Hum Biol* 66:639–668.
- Barbujani G. 1997. DNA variation and language affinities. *Am J Hum Genet* 61:1011–1014.
- Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. 1997. An apportionment of human DNA diversity. *Proc Natl Acad Sci USA* 94:4516–4519.
- Barrai I, Rosito A, Cappellozza G, Cristofori G, Vullo C, Scapoli C, Barbujani G. 1984. Beta-thalassemia in the Po Delta: selection, geography, and population structure. *Am J Hum Genet* 36:1121–1134.
- Bertorelle G, Excoffier L. 1998. Inferring admixture proportions from molecular data. *Mol Biol Evol* 15:1298–1311.
- Calafell F, Underhill P, Tolun A, Anglicheva D, Kalaydjieva L. 1996. From Asia to Europe: mitochondrial DNA sequence variability in Bulgarian and Turks. *Ann Hum Genet* 60:35–49.
- Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci USA* 85:6002–6006.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton, NJ: Princeton University Press.
- Chakraborty R, Kamboh M, Nwankwo M, Ferrell RE. 1992. Caucasian genes in American Blacks: new data. *Am J Hum Genet* 50:145–155.
- Chen HW, Lowther D, Avramopoulos D, Antonarakis SE. 1994. Homologous loci DXY156X and DXY156Y contain a polymorphic pentanucleotide repeat (TAAAA)_n and map to human X and Y chromosomes. *Hum Mutat* 4:208–211.
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bertranpetit J. 1996. Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 13:1067–1077.
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J. 1995. Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 63:1824–1838.
- Deka R, Jin L, Shriver MD, Yu Ling M, Decroo S, Hundrieser J, Bunker C, Ferrell RE, Chakraborty R. 1995. Population genetics of dinucleotide (dC-dA)_n(dG-dT)_n polymorphism in world populations. *Am J Hum Genet* 56:461–174.
- Endress G. 1988. An introduction to Islam. New York: Columbia University Press.
- Excoffier L, Harding RM, Sokal RR, Pellegrini B, Sanchez-Mazas A. 1991. Spatial differentiation of RH and GM haplotype frequencies in sub-Saharan Africa and its relation to linguistic affinities. *Hum Biol* 63:273–307.
- Excoffier L, Smouse P, 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. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
- Guglielmino CR, Piazza A, Menozzi P, Cavalli-Sforza LL. 1990. Uralic genes in Europe. *Am J Phys Anthropol* 83:57–68.
- Hammer MF, Horai S. 1995. Y chromosome DNA variation and the peopling of Japan. *Am J Hum Genet* 56:951–962.
- Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am J Hum Genet* 55:175–189.
- Harding RM, Sokal RR. 1988. Classification of the European language families by genetic distance. *Proc Natl Acad Sci USA* 85:9370–9372.
- Jorde LB, Watkins WS, Bamshad MJ, Dixon 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.
- Kaiser MP, De Knijff P, Dieltjed P, Krawczak M, Nagy M, Zerjal T, Pandya A, Tyler-Smith C, Roewer L. 1997. Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18:1602–1607.
- Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321–1334.
- Long JC. 1991. The genetic structure of admixed populations. *Genetics* 127:417–428.
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonn -Tamir B, Sykes B, Torroni A. 1999. The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232–249.
- Maniatis T, Fitch EF, Sambrook J. 1982. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press.
- Morton NE, Kenett R, Yee S, Lew R. 1982. Bioassay of kinship in populations of Middle Eastern origin and controls. *Curr Anthropol* 23:157–167.

- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- Nei M, Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97:145–163.
- Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J. 1999. Sex-specific migration patterns in Central Asian populations, revealed by the analysis of Y-chromosome short tandem repeats and mtDNA. *Am J Hum Genet* 65:208–219.
- Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L. 1997. Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet* 61:1015–1035.
- Rando JC, Pinto F, Gonz ales AM, Hern andez M, Larruga JM, Cabrera VM, Bandelt H-J. 1998. Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Am J Hum Genet* 62:531–550.
- Renfrew C. 1987. Archaeology and language. The puzzle of Indo-European origins. London: Jonathan Cape.
- Renfrew C. 1989. Models of change in language and archaeology. *Trans Philol Soc* 87:103–155.
- Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. 1998. Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* 62:241–260.
- Roux JP. 1984. Histoire des Turcs. Paris: Librairie Arth me Fayard.
- Ruhlen M. 1991. A guide to the world's languages. Volume 1: classification, 2nd ed. London: Edward Arnold.
- Sajantila A, P  bo S. 1995. Language replacement in Scandinavia. *Nat Genet* 11:359–360.
- Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T, Issel-Tarver L, Di Rienzo A, P  bo S. 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 5:42–52.
- Salem AH, Badr FM, Gaballah MF, P  bo S. 1996. The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai peninsula. *Am J Hum Genet* 59:741–743.
- Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, Torroni A, Modiano D, Wallace DC, Kidd KK, Olckers A, Moral P, Terrenato L, Akar N, Qamar R, Mansoor A, Mehdi SQ, Meloni G, Vona G, Cole DEC, Cai W, Novelletto A. 1997. Differential structuring of human populations for homologous X and Y microsatellite loci. *Am J Hum Genet* 61:719–733.
- Seielstad M, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278–280.
- Semino O, Passarino G, Oefner PJ, Lin AL, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill P. 2000. The genetic legacy of Paleolithic *Homo sapiens*: a Y chromosome perspective. *Science* 290:1155–1159.
- Simoni L, Gueresi P, Pettener D, Barbuji G. 1999. Patterns of gene flow inferred from genetic distances in the Mediterranean region. *Hum Biol* 71:399–415.
- Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbuji G. 2000. Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet* 66:262–278.
- Sokal RR. 1988. Genetic, geographic, and linguistic distances in Europe. *Proc Natl Acad Sci USA* 85:1722–1726.
- Sokal RR, Oden NL, Thomson BA. 1988. Genetic change across language boundaries in Europe. *Am J Phys Anthropol* 76:337–361.
- Sokal RR, Oden NL, Legendre P, Fortin M-J, Kim J, Thomson BA, Vaudor A, Harding RM, Barbuji G. 1990. Genetics and language in European populations. *Am Natur* 135:157–175.
- Sokal RR, Oden NL, Walker J, Di Giovanni D, Thomson BA. 1996. Historical population movements in Europe influence genetic relationships in modern samples. *Hum Biol* 68:873–898.
- Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC. 1998. MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am J Hum Genet* 63:1473–1491.
- Torroni A, Schurr T, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mitochondrial DNAs. *Am J Hum Genet* 53:591–608.
- Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci USA* 86:9350–9354.
- Walsh PS, Metzger DA, Higuchi R. 1991. Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10:298–317.
- Ward RH, Reed A, Valencia D, Frazier V, P  bo S. 1993. Genetic and linguistic differentiation in the Americas. *Proc Natl Acad Sci USA* 90:10663–10667.
- Wright S. 1969. Evolution and the genetics of populations. Volume II: the theory of gene frequencies. Chicago: Chicago University Press.