Am. J. Hum. Genet. 63:1824–1838, 1998

Trading Genes along the Silk Road: mtDNA Sequences and the Origin of Central Asian Populations

David Comas,[1,2,*] Francesc Calafell,[1,3,*] Eva Mateu,[1] Anna Pérez-Lezaun,[1,4] Elena Bosch,[1] Rosa Martínez-Arias,[1] Jordi Clarimon,[1] Fiorenzo Facchini,[5] Giovanni Fiori,[5] Donata Luiselli,[5] Davide Pettener,[5] and Jaume Bertranpetit[1]

[1]Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona; [2]Department of Forensic Medicine, University of Helsinki, Helsinki; [3]Department of Genetics, Yale University School of Medicine, New Haven; [4]Center for Human Genetics, Edith Cowan University, Joondalup, Western Australia; and [5]Dipartimento di Biologia e.s., Unità di Antropologia, Università di Bologna, Bologna

Summary

Central Asia is a vast region at the crossroads of different habitats, cultures, and trade routes. Little is known about the genetics and the history of the population of this region. We present the analysis of mtDNA control-region sequences in samples of the Kazakh, the Uighurs, the lowland Kirghiz, and the highland Kirghiz, which we have used to address both the population history of the region and the possible selective pressures that high altitude has on mtDNA genes. Central Asian mtDNA sequences present features intermediate between European and eastern Asian sequences, in several parameters—such as the frequencies of certain nucleotides, the levels of nucleotide diversity, mean pairwise differences, and genetic distances. Several hypotheses could explain the intermediate position of central Asia between Europe and eastern Asia, but the most plausible would involve extensive levels of admixture between Europeans and eastern Asians in central Asia, possibly enhanced during the Silk Road trade and clearly after the eastern and western Eurasian human groups had diverged. Lowland and highland Kirghiz mtDNA sequences are very similar, and the analysis of molecular variance has revealed that the fraction of mitochondrial genetic variance due to altitude is not significantly different from zero. Thus, it seems unlikely that altitude has exerted a major selective pressure on mitochondrial genes in central Asian populations.

Received December 12, 1997; accepted for publication September 25, 1998; electronically published December 2, 1998.

Address for correspondence and reprints: Dr. Jaume Bertranpetit, Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Catalonia, Spain. E-mail: jaumeb.bertranpetit@cexs.upf.es

* These two authors contributed equally to this work.

0002-9297/98/6306-0029$02.00

Introduction

Central Asia is a vast territory very poorly known genetically, although it has played a crucial role in the history of humankind and has been an area where cultural and linguistic changes are known to have been of key importance. It is often regarded as a borderland between East and West, without a unique, particular history, which is not the case (Bowles 1977; Sellier and Sellier 1993).

Central Asia, as defined by Soviet scholars, encompasses the territories east of the Caspian Sea to the current boundaries of China along the Pamir, the Hindu Kush and farther to the northeast, and it comprises the republics of Uzbekistan, Tajikistan, Turkmenistan, Kirghizstan, and part of Kazakhstan; in Western literature, Mongolia, Tibet, and Sinkiang (pinyin Xinjiang, western China) sometimes are included. It is a territory of vast contrasts, with most of the land occupied by high-altitude tracts or vast cold deserts, both unfavorable for large and stable human settlements. However, the river basins have been occupied since early times, and the steppes have offered a good land for an itinerant pastoral economy.

A territory located at the edge of the western Asian empires, crossed by the Silk Road, with long-lasting contacts with India yet open to the steppes of the north, it is a case in point when one is trying to understand the genetic consequences of complex cultural phenomena such as acculturation, assimilation, and syncretism; overlapping of economies, languages, and ways of life; and migrations, expansions, and conquests.

The role of central Asia in early human evolution and history is not well established. According to an old, long-dismissed hypothesis, the nearby Altai region could have been the origin of humankind. It is known that the region was populated during the lower Paleolithic, and there is ample evidence of settlement during the middle Paleolithic, including Teshik-Tash, the easternmost site from which Neanderthal remains have been recovered.

It is not clear, however, whether the region was part of a "maturation" phase of anatomically modern humans, a thruway in the colonization of Europe and eastern Asia, or a place where Asian and European groups met after their expansion (Bowles 1977).

The advent of the Neolithic seems to have been a local development without significant external population inputs. The domestication of the horse in the steppes (Anthony and Brown 1991) and, subsequently, the development of wheeled vehicles (Anthony and Vinogradov 1995) had a major impact on world history, as mobility increased dramatically and warfare was profoundly changed. The human history of central Asia has been determined by the crudeness of the physical geography, which has not allowed stable human settlements. The Scythians are the first people in the historic record of central Asia, around the 7th century B.C., and they are described as having European morphological traits, both by ancient Chinese texts and the Greek historian Herodotus in the 5th century B.C. In the 2d century B.C., the Chinese established a trade route from the Mediterranean Sea to eastern Asia (the Silk Road), which connected the East and the West of the continent during >16 centuries, until it was replaced by safer ocean trade routes. During the 3d and 4th centuries A.D., Turkic hordes of Siberian origin replaced the Indo-European peoples in central Asia and created a great empire from Mongolia to the Black Sea, revived by Genghis Khan during the 13th century. Later, the Chinese and Russian empires established their rule over the vast territories of central Asia.

Genetically, central Asia is one of the least-studied major regions of the world. The analysis of classical genetic markers has been done mainly by Soviet scholars, and, in the global revision by Cavalli-Sforza et al. (1994), little is added to the idea that central Asia is intermediate between Asia and Europe.

We have sequenced mtDNA hypervariable region I in samples of the Uighurs, the Kazakhs, and the Kirghiz, three ethnic groups of central Asia, in order to investigate the origins and evolution of central Asian populations and to address several hypotheses concerning the population history of Eurasia. Furthermore, we have collected samples from high- and low-altitude populations and have investigated the possibility that selective pressures on mtDNA genes have been caused by low oxygen pressure.

Material and Methods

Population Samples

A 360-nucleotide sequence in hypervariable region I of the mitochondrial control region (positions 16024–16383 in the reference sequence; Anderson et al. 1981) was analyzed in 55 Uighurs, 55 Kazakhs, 47 Kirghiz from Sary-Tash, and 48 Kirghiz from Talas (fig. 1). Each sample comprises ethnically homogeneous, autochthonous, unrelated, healthy male donors, from whom appropriate informed consent was obtained. The Uighurs were sampled in the village of Penjim (600 m above sea level) in the Panfilov district, Taldy-Corgan region, in the easternmost section of Kazakhstan, only 18 km from the boundary with China. This region is inhabited mostly by Uighurs, who emigrated from Sinkiang (pinyin Xinjiang; Chinese Uighur autonomous region) during recent decades. The Kazakh samples were collected in the villages of the Kegen valley (2,100 m above sea level; Almaty region, Kazakhstan), a high plain a few kilometers from the northern slopes of the Tien Shan range. Two Kirghiz samples, one each from Sary-Tash in the Pamir and from Bakai Ata in the Talas Valley, were analyzed. Sary-Tash is an isolated high-altitude village (3,200 m above sea level) a few kilometers from the border with Tajikistan and China, in the heartland of the Pamir mountains. The village of Bakai Ata (900 m above sea level) is located in the Talas valley in the northernmost section of Kirghizstan, along the ancient Silk Road trade routes to Kazakhstan and Uzbekistan.

Blood samples were collected by the Italian team (F.F., G.F., D.L., and D.P.) within the CAHAP (Central Asia High Altitude People) research program, which is promoted jointly by the Laboratory of Anthropology of the Academy of Science of Kazakhstan and the Anthropology Unit of the University of Bologna. The objective of CAHAP is the study of (*a*) human adaptability to high altitude (Pettener et al. 1997), (*b*) body composition (Battistini et al. 1995; Bedogni et al. 1997; Facchini et al. 1998), and (*c*) genetic variability (Facchini et al. 1997; Pettener et al. 1997), in central Asian mountain popu-

Figure 1 Geographic map of sampled area. The four central Asian samples are in boldface. KIR-TALAS = Kirghiz from Talas, and KIR-SARYTASH = Kirghiz from Sary-Tash. The Kazakhs and the Sary-Tash Kirghiz live at elevations >2,000 m above sea level.

lations. Two Italo-Kazakh expeditions were undertaken—one in 1993 in Kazakhstan and the other in 1994 in Kirghizstan—in order to collect new information on these topics.

mtDNA Extraction and Sequencing

Genomic DNA was extracted from whole blood by a phenol-chloroform extraction method after a digestion with proteinase K. The control-region I sequence was PCR-amplified with primers L15996 and H16401 (Vigilant et al. 1989). The sequence reaction was performed on each strand by means of a DNA sequencing kit, Dye Terminator Cycle Sequencing with Ampli*Taq* DNA Polymerase (Perkin-Elmer). The product of the sequence reaction was run in an ABI PRISM 377 (Perkin-Elmer) automatic sequencer, and the sequences were aligned by the ESEE computer program (Cabot 1988). One sib of one individual was also sequenced as a control. The sequences can be requested from F. C. by E-mail (francesc.calafell@cexs.upf.es).

Statistical Analysis

Nucleotide diversity (Nei and Tajima 1981) was estimated as $(n/n-1)(1/L)\Sigma_{j=1}^{L}(1-\Sigma_{i=1}^{4}x_{ij}^{2})$, where n is sample size, L is sequence length, and x_{ij} is the frequency of the ith nucleotide (A, C, G, or T) at position j. Similarly, sequence diversity was estimated as $(n/n-1)\Sigma_{i=1}^{k}(1-p_i^2)$, where p_i is the frequency of each of the k different sequences in the sample. The significance of the sequence-diversity difference between two of the central Asian populations was tested by means of nonparametric permutation procedure: the empirical null distribution of the difference in sequence diversity was obtained by permuting the individual sequences at random across populations and repeating the permutation procedure 5,000 times. The probability of finding, between two populations, a sequence-diversity difference larger than that actually observed was estimated on the basis of the empirical distribution.

Genetic homogeneity among populations was tested by analysis of molecular variance (AMOVA; Excoffier et al. 1992), by means of the Arlequin package (Schneider et al. 1996). For most other calculations, standard packages such as PHYLIP 3.5c (Felsenstein 1989) were used, and some programs (available, on request, from F.C., at francesc.calafell@cexs.upf.es) were written specifically for this study.

For comparison, data from the following, other populations were used: Turks (Calafell et al. 1996; Comas et al. 1996; Richards et al. 1996), Chinese (nonaboriginal Taiwanese, presumably of Han descent; Horai et al. 1996), Ainu (Horai et al. 1996), Koreans (Horai et al. 1996), Altai from the Gorno-Altai region (Siberia; Shields et al. 1993), Mongolians (Kolman et al. 1996), Havik (Mountain et al. 1995), British (Piercy et al. 1993), and Middle Easterners (Di Rienzo and Wilson 1991). An extended database was used in some analyses; the additional sequences were obtained from African samples (Vigilant et al. 1991; Graven et al. 1995; Watson et al. 1996, 1997), European samples (Di Rienzo and Wilson 1991; Handt et al. 1994 [and additional Austrian sequences in the database described in Handt et al. 1998]; Pult et al. 1994; Bertranpetit et al. 1995; Sajantila et al. 1995; Calafell et al. 1996; Côrte-Real et al. 1996; Francalacci et al. 1996; Pinto et al. 1996; Richards et al. 1996; Stenico et al. 1996), Indian samples (Mountain et al. 1995), Polynesian samples (Lum et al. 1994), and Native American samples (Ward et al. 1991, 1993, 1996; Shields et al. 1993; Santos et al. 1994; Batista et al. 1995; Kolman et al. 1995. Genetic distance between populations was estimated by the expression $D = d_{ij} - [(d_i + d_j)/2]$, where d_{ij} is the raw mean nucleotide pairwise difference between populations i and j, and where d_i and d_j are the raw mean nucleotide pairwise differences within populations i and j, respectively (Rao 1982; Nei 1987). When negative genetic distances were found, a small constant was added to the entire matrix to make it positive; this procedure does not alter the genetic relationships and enabled us to use tree-building algorithms. Neighbor-joining trees (Saitou and Nei 1987) were built from the distance matrix, and the bootstrap method (Efron 1982) was used to estimate the robustness of the branches. The distance matrix was also represented by means of principal-coordinate analysis (Gower 1966; Cavalli-Sforza et al. 1994, p. 42)

Results

Sequence Diversity

A total of 146 different sequences (see fig. A1, in the Appendix) defined by 108 variable nucleotide positions were obtained from positions 16024–16383 in the reference sequence (Anderson et al. 1981), from the 205 individuals analyzed. A set of descriptive parameters of the sequences of each population are shown in table 1. All observed polymorphisms were nucleotide substitutions, except for that in one individual from Sary-Tash who presented a deletion of one C in the run of three C's at positions 16294–16296. All variable positions found in these central Asian samples have been described, elsewhere, in other populations (see the description of the extended population database, in the Material and Methods section).

Central Asian populations presented sequence diversities of .984–.995 (table 1), with the lower values corresponding to the two high-altitude populations. European populations exhibited mtDNA sequence diversities ranging from .936 (Danes) to .984 (Bavarians); the range in eastern Asia is from .947 (Ainu) to .993 (Chinese). Sequence diversity reaches .988 in Turks, .995 in

Table 1

Diversity Parameters for Central Asian Populations

	Kazakh	Kirghiz (Sary-Tash)	Kirghiz (Talas)	Uighur
No. of individuals	55	47	48	55
No. of different sequences	45	34	43	46
No. of variable positions	64	59	56	63
Sequence diversity	.990	.984	.995	.993
Ewens's θ	113.33	53.72	194.50	129.71
Nucleotide diversity	.0185	.0170	.0180	.0164
Mean pairwise differences (SD)	6.64 (2.94)	6.12 (2.93)	6.48 (2.59)	5.91 (2.52)
Tajima's D	−1.814	−1.902	−1.699	−1.969

Middle Easterners, and .955 in Mongols. Sequence diversity in central Asia is, therefore, among the highest in Eurasia.

To test whether the reduction in sequence diversity in high-altitude populations is statistically significant, a permutation test was performed, as described in the Material and Methods section. The most extreme difference was between the two Kirghiz samples: Sary-Tash, at an altitude of 3,200 m above sea level, and Talas, at an altitude of 900 m above sea level; that difference was found to be close to the significance level ($P = .066$). The other five possible pairs of populations were also tested; in none of them was the difference significant.

The parameter θ, from the equation $\theta = 2N_e\mu$, was estimated as suggested by Ewens (1972); N_e represents the effective population size, and μ is the mutation rate; therefore, interpopulation differences in θ should be a function of effective population size, since mutation rate is presumably constant across populations. High-altitude Kirghiz (Sary-Tash) present a θ value lower than those in the rest of the central Asian samples (table 1), which could be due to a number of causes, such as (i) a smaller exposure to external gene flow, (ii) long-term smaller effective population size, (iii) a bottleneck in the Kirghiz settlement in the high altitudes of the Pamir mountains, (iv) greater variance in the number of offspring per mother in the Kirghiz, and/or (v) selection. A combination of the demographic factors (i.e., i–iv), linked to the colonization of an isolated, ecologically poor habitat, seems more plausible; as discussed below, selection is not likely to have operated extensively in high-altitude mtDNA sequences in central Asia.

Sequence Sharing

All four central Asian populations share at least two sequences: the reference sequence and the sequence that presents both a T at position 16223 and a C at position 16362. Seventeen of the 146 different sequences found were shared by at least two central Asian samples. When those populations are compared to some sub-Saharan African populations in our extended database, central Asian populations are seen to share only two sequences with Turkana (one is presented by the Kazakhs, and the other is presented by the Sary-Tash Kirghiz), one Somali sequence is found in the Uighurs, and the reference sequence is found in the Tuareg. Few central Asian sequences are found in America—only one in the Pacific Northwest Haida, one in the Panamanian Kuna, and two in Circumarctic populations. Central Asian populations share a few sequences with eastern Asian groups; for instance, six sequences are shared with the Han Chinese, five are shared with Koreans, and four are shared with the Ainu. Central Asian populations share a total of 15 sequences with Mongols (Kolman et al. 1996) and 9 sequences with Turks (96 individuals, pooled from Calafell et al. 1996; Comas et al. 1996; Richards et al. 1996). Six of the nine sequences shared with Turks, which characterize 18 Turkish individuals, are also found all over Europe, whereas the rest, found only once each in Turks, are common in eastern Asian populations but have not been described in Europeans. Central Asian populations share a total of 17 different sequences with European populations (in a combined set of 1,420 European individuals) and a total of 12 with eastern Asian populations. Of the sequences shared with Europeans or with East Asians, only two sequences have been found in Europe as well as in eastern Asia. It should be stressed that, although central Asians share sequences with Europeans and eastern Asians, these two latter groups of populations have very few sequences in common. Only one Chinese, one Ainu, and two Korean sequences have been found in Europe, which implies that Europeans and eastern Asians present nearly nonoverlapping mtDNA sequence pools.

Nucleotide Diversity

For every site, the nucleotide in the reference sequence is the most frequent in each population, except for position 16223 which presents a C in the reference but shows a high frequency of T in the populations studied (50.9% in Kazakhs, 48.9% in the Sary-Tash Kirghiz, 68.8% in the Talas Kirghiz, and 41.8% in the Uighurs). A high level of polymorphism was also found at position 16362, which presents a T in the reference sequence and a high frequency of C in the populations studied (32.7% in Kazakhs, 29.8% in the Sary-Tash Kirghiz, 43.8% in

the Talas Kirghiz, and 32.7% in Uighurs). Nucleotide diversity in central Asia ranges from .0164 in the Uighurs to .0185 in the Kazakhs. Those values are similar to those in Mongolians (.0180) and slightly smaller than those in eastern Asian populations, which present nucleotide diversities from .0173 in Koreans to .0195 in the Ainu. European populations present nucleotide diversities that are less than those found in central Asia; they range from .0082 in Basques (pooled from data reported by Bertranpetit et al. [1995] and Côrte-Real et al. [1996]) to .0140 in Tuscany. Turks present a nucleotide diversity of .0155, smaller than that of central Asian populations, and Middle Easterners present the highest value (.0197) observed in European and Asian populations. Therefore, nucleotide polymorphism in mtDNA sequences in central Asian populations is intermediate between those reported for Europe and those reported for eastern Asia. We cannot exclude, though, that such values are due to the effect of one or a few nucleotide positions. If the most polymorphic position (i.e., position 16223) is excluded from the analysis, nucleotide diversities become .0151–.0171 in central Asia, .0166–.0185 in eastern Asia, .0186 in the Middle East, .0142 in Turkey, and .0081–.0133 in Europe. Although, when position 16223 is removed from the analysis, the absolute values of nucleotide diversity in central Asia decrease and approach those found in Turkey, both the general trend of nucleotide diversity in Eurasia and the relative position of central Asia in it remain. This is also true when the second most polymorphic position, position 16362, is also removed: with both position 16223 and position 16362 excluded, nucleotide diversities are .0139–.0159 in central Asia, .0152–.0173 in eastern Asia, .0180 in the Middle East, .0139 in Turkey, and .0079–.0131 in Europe.

Pairwise Differences

The mean pairwise differences in central Asian populations (table 1) range from 5.91 in the Uighurs to 6.64 in the Kazakhs. When these values are compared with those in other populations, it is apparent that, on one hand, central Asian populations have slightly lower values than the eastern Asian groups, whose mean values range from 6.68 in Koreans to 7.51 in the Ainu, and, on the other hand, the values found in central Asia are higher than the mean pairwise differences found in Europeans (range 3.15–5.03; Comas et al. 1997) and Turks (5.45). Again, our results show an intermediate position of central Asia, between Europe and eastern Asia.

Pairwise-difference distributions of the four central Asia populations are clearly bell-shaped, with peaks at five (Uighur), six (Kazakh and Sary-Tash Kirghiz), and seven (Talas Kirghiz) differences. The distribution derived by Rogers and Harpending (1992) for pairwise differences could be fitted to the distributions observed in all four central Asian populations. All four central Asian samples presented low, negative Tajima's (1989) *D* statistics, and, except for the Talas Kirghiz, those *D* statistics were significantly lower than the values that would be expected under equilibrium. Even with mutation rates varying across nucleotide positions, low Tajima's *D* statistics and bell-shaped pairwise-difference distributions can be interpreted as the hallmark of an ancient population expansion (Aris-Brosou and Excoffier 1996). However, given the uncertainty about the mutation-rate estimates for the hypervariable region of human mtDNA (Howell et al. 1996; Parsons et al. 1997), there is some controversy about which expansion event is reflected in human mtDNA (Pääbo 1996; Howell and Mackey 1997; Macaulay et al. 1997). Thus, Harpending et al. (1993) point to the Pleistocene expansion (~100,000 years ago) as being linked to the spread of anatomically modern humans out of Africa, whereas Watson et al. (1996) favor a later date (10,000–2,000 years ago) and associate bell-shaped pairwise distributions with populations that have undergone the ecological and demographic transition from food gathering to food production. Nonetheless, it seems difficult to accept such recent dates, even in light of high mutation rates (Bandelt and Foster 1997).

Genetic Structure of Central Asian Populations

The genetic structure of the four central Asian samples was investigated by AMOVA (Excoffier et al. 1992). When the four populations were treated as single groups, we found that 99.54% of the genetic variation remained within populations, whereas the remaining 0.46% (which is not significantly different from 0 [$P = .0949$; 10,000 iterations]) of the variance could be attributed to differences between populations. Next, we subdivided the populations into mountain dwellers (the Kazakhs and Sary-Tash Kirghiz, both from localities at an altitude of 2,000 m above sea level) and plains dwellers (the Uighurs and Talas Kirghiz). The fraction of genetic variance that could be apportioned to this partition was, according to AMOVA, −0.52% ($P = .6635$), indicating that genetic variance within any ecological group was larger than that between groups and, therefore, that the division by altitude is not reflected in mtDNA variation. Similarly, when populations were grouped by language (three groups—Kirghiz, Kazakh, and Uighur, all belonging to the Turkic branch of the Altaic family), linguistic affiliation did not contribute to the genetic variance (−0.47%; $P = .8325$). In summary, mtDNA sequence variation in these central Asian samples appears to be homogeneous among populations, with (within the statistical power of AMOVA) no discernible

differentiation according to altitudinal habitat or linguistic group.

Genetic Distances

Genetic distances (not shown) between central Asian populations and eastern Asian, Indian, western Asian, and European populations were calculated as described in the Material and Methods section. The four central Asian populations presented the shortest genetic distances among themselves, and Mongolians are the population genetically closest to these central Asian groups. If African populations are added to the analysis, they present large genetic distances to all other populations. It is interesting to note that Turks present shorter genetic distances to the British than to central Asians, even though the central Asian populations' samples in the present study speak Turkic languages. A neighbor-joining tree was built as described in the Material and Methods section, with the genetic distances estimated on the basis of the mismatch-intermatch distance. The robustness of the tree was assessed by means of 1,000 bootstrap replicates (Efron 1982; Felsenstein 1985), a consensus tree was built, and bootstrap supports >50% have been represented on its nodes (fig. 2). It is evident that central Asian populations occupy a position intermediate between the eastern Asian (Chinese, Korean, and Ainu) and the Western (Middle Eastern, British, and Turk) populations. Not surprisingly, the most robust nodes are those that cluster together the eastern Asian populations on one end and the Western populations on the other end. No robust branches subdivided the central Asian populations.

Since a tree representation of the distance matrix may be misread as a succession of splits, we performed a principal-coordinate analysis on the distance matrix. The first two principal coordinates are represented in figure 3. The first principal coordinate explained 51.9% of the variance, and the second principal coordinate explained an additional 21.1% (73.0% combined). The first principal coordinate placed the populations in an east-west cline similar to that displayed by the neighbor-joining tree; most Asian populations presented intermediate values for the second principal coordinate, and only the Ainu and the Altai, both of which are small, isolated populations, presented extreme and opposed coordinates.

Admixture Analysis

The evidence presented above seems compatible with an admixture scenario for the origin of the central Asian mtDNA sequence pool, as will be discussed in detail below. As a first, rough approach to the measurement

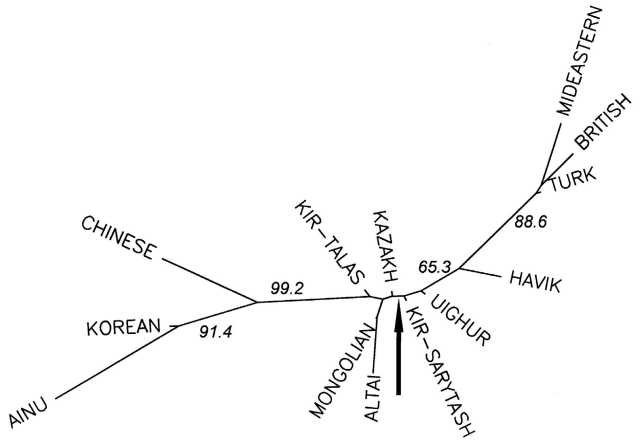

Figure 2 Neighbor-joining tree of several European and Asian populations. For the references to the original data, see the Material and Methods section. Bootstraps supports >50% are shown in the nodes of the tree. The arrow points to the segment from which an African outgroup (!Kung San) would branch.

of admixture, one can compute an admixture proportion from the genetic-distance matrix by the triangle method, as described by Cavalli-Sforza et al. (1994, p. 57). The proportion m of European sequences found by this method is $m = .49$ for the Kazakhs, $m = .52$ for the highland Kirghiz, $m = .36$ for the lowland Kirghiz, and $m = .55$ for the Uighurs. Overall, the proportions of European and eastern Asian sequences in central Asian sequences would be .48 and .52, respectively. However, we can refine those estimates by trying to ascertain the possible European origin versus eastern Asian origin of each of the sequences that we found in central Asia, by a phylogenetic approach. To that effect, we constructed a tree (fig. 4) with sequences from Europe (British and Tuscan samples; 104 different sequences) and eastern Asia (Ainu, Chinese, and Korean samples; 113 different sequences). Only one sequence was shared between Europeans and Asians. We could identify 16 groups (table 2), which encompassed 82.5% of the sequences. Only one group contained significant numbers of both European and eastern Asian sequences; another group contained 1 European sequence and 25 eastern Asian sequences and was considered an Asian group. The same groups were found also in a median network constructed with the same sequences (results not shown; Bandelt et al. 1995), and almost all match the most ancient nodes in the European and Asian phylogenies of mtDNA sequences (H.-J. Bandelt, personal communication). If a central Asian sequence presented the nucleotide changes defining that group, it was assigned to that group and, consequently, to Europe or eastern Asia. When sequences did not present any of the groups of defining positions of each lineage (table 2), we compared them

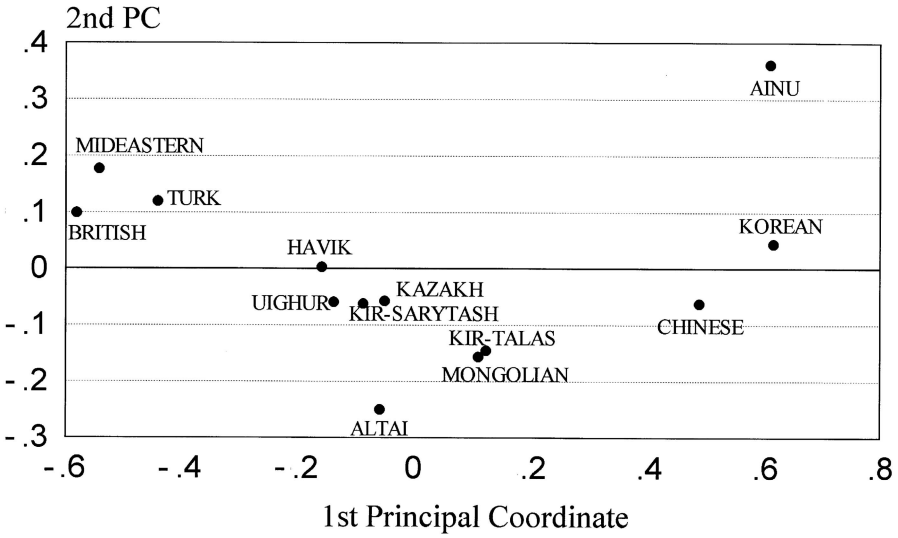

Figure 3 Principal-coordinate analysis of distance matrix, among Asian and European populations. The first principal coordinate explains 51.9% of the variation, and the second explains 21.1%, for a combined percentage of 73.0%. Population abbreviations are as in figure 1.

with our extended database. For a sequence to be assigned either to eastern Asia or to Europe, it had to be identical to, or to differ in no more than two nucleotides from, a sequence found exclusively in Europe or eastern Asia. That approach allowed us to identify 93.7% of central Asian sequences as belonging to an already sequenced eastern Asian or European lineage (table 3). An average of 33.2% of the individuals in our central Asian samples bore a sequence belonging to a European lineage. This fraction became 35.4% when the unassigned sequences were not taken into account. The proportion of eastern Asian, European, and unassigned sequences was not significantly different across central Asian populations ($\chi^2 = 7.67$, 6 df, $P = .264$).

Discussion

Torroni et al. (1994) did not find, in mtDNA RFLPs in Tibetans, any selective effects attributable to high altitude; we can reach similar conclusions with control-region sequences when comparing lowland and highland central Asian populations. As stated above, the genetic variance attributable to altitude is not significantly different from 0. Moreover, Fu's (1997) F_s-test presents very similar values in the highland and the lowland Kirghiz (-20.249 and -20.627, respectively, both of which are significantly negative [$P < .005$]). Such an excess of low-frequency segregating nucleotides could be produced by a population expansion, selection, or both. It is unlikely that selection would have acted to almost exactly the same extent that a population expansion did in lowland populations, as would be indicated by Fu's F_s, and that, after selection, mutation would have regenerated, at high altitudes, a sequence pool that resembles so closely that of low-altitude populations. Therefore, we are unable to detect any effects of selection in the mtDNA control region (and, given the absence of recombination in mtDNA, that is true for the whole molecule). Nevertheless, this does not preclude the existence of selective changes in one or more of the multiple nuclear genes coding for respiratory-chain proteins.

The results of the present study consistently show that the central Asia mtDNA sequences present features that are intermediate between those found in Europe and eastern Asia. This is especially patent in the following: (i) the cline of the frequency of certain nucleotides in specific positions, such as those found at positions 16223 and 16362; (ii) polymorphism at the nucleotide level, as measured by nucleotide diversity—even when the effects of clinal nucleotide positions are discounted; (iii) the average pairwise-difference values, which are intermediate between those of Europe and those of eastern Asia; and (iv) genetic distances, which locate the central Asian populations between Europe and eastern Asia. Several population history scenarios could have produced the intermediate genetic features of central Asian mtDNA sequences; some hypotheses that could be put forth—such as an Asian colonization of Europe, or vice versa—find no support in archaeological knowledge and would contradict other mtDNA evidence (Ballinger et al. 1992). However, some of the analyses that we performed will allow us to assess the degree to which other, more plausible hypotheses are supported by mtDNA evidence.

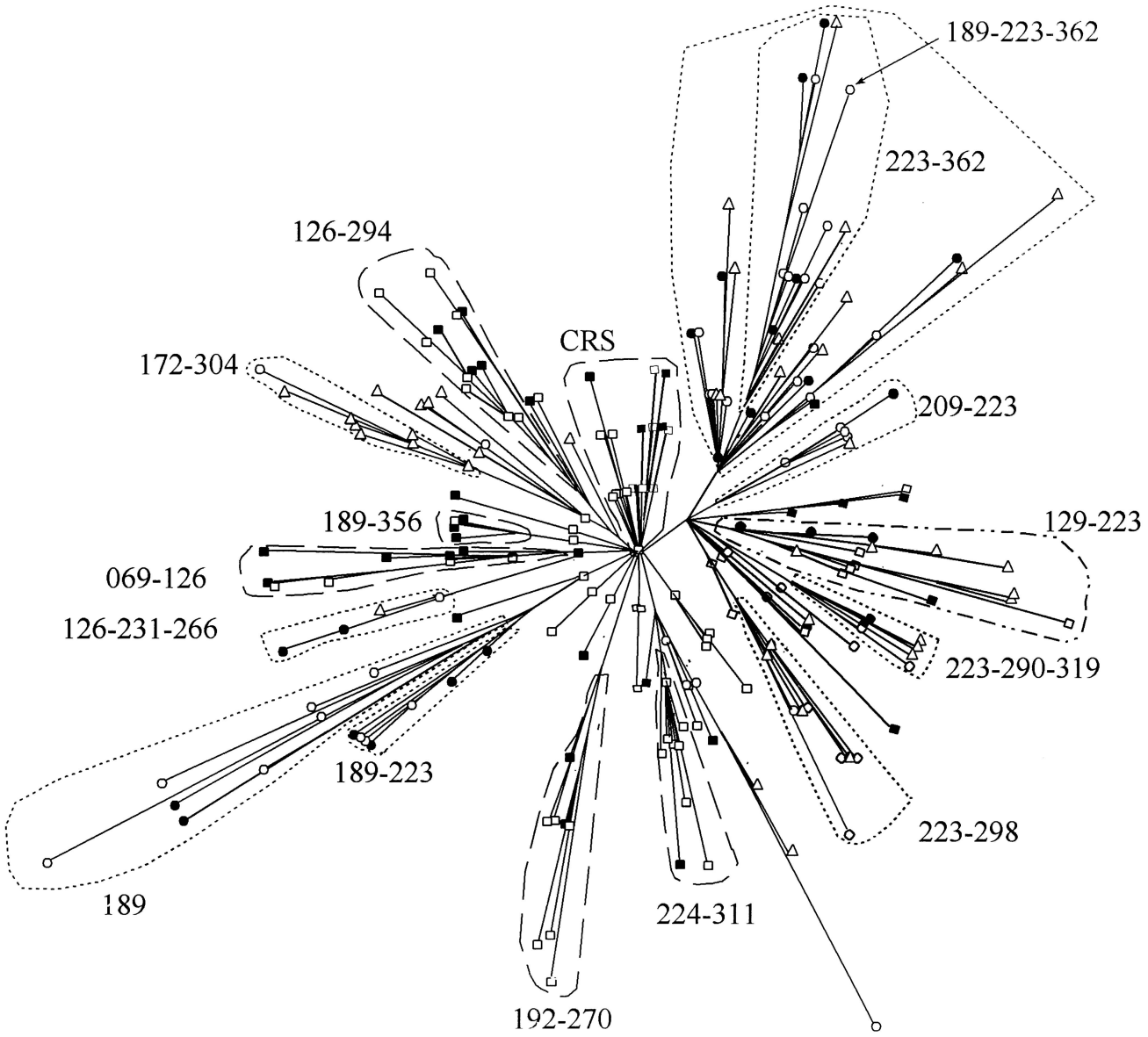

Figure 4 Neighbor-joining tree of eastern Asian (Chinese, Ainu, and Korean) and European (British and Tuscan) sequences. Genetic distances between sequences were estimated via the Kimura two-parameter model, with the transition:transversion ratio set to 15:1. The tree was postprocessed by rounding branch lengths to whole numbers, to eliminate ghost links (Richards et al. 1996). Dotted lines encircle the major lineages (table 2), and figures represent their defining positions (add 16,000 to obtain the numbering used by Anderson et al. [1981]). "CRS" denotes the Cambridge Reference Sequence (Anderson et al. 1981) and those sequences directly derived from it. Circles and triangles denote sequences found in eastern Asia (unblackened circles denote Korean sequences; blackened circles denote Ainu sequences, and triangles denote Chinese sequences), and squares denote sequences found in Europe (unblackened squares denote British sequences, and unblackened squares denote Tuscan sequences).

Hypothesis A. Europe and Eastern Asia: Colonized from Central Asia

A possible scenario for the colonization of Europe and Asia by anatomically modern humans would involve an initial "maturation phase" in central Asia. Under a replacement model of modern human origins, it is often assumed that the migratory pathways leading to Europe and Asia from Africa split somewhere in central Asia or in neighboring regions. With the exception of the Middle

Table 2

Sequence Groups Identified in a Sequence Tree Built with 113 Eastern Asian (Han Chinese, Korean, and Ainu) and 104 European (British and Tuscan) Sequences

Group-Defining Position(s)	No. of Sequences in	
	Eastern Asian	European
16189C	8	0
16189C/16223T	7	0
16126C/16231C/16266T	4	0
16172C/16304C	9	0
16223T/16298C	9	0
16223T/16290T/16319A	7	0
16223T/16209C	6	0
16223T/16362C	25	1
16223C/16362C/16189 C	8	0
16223T/16129A	9	6
16126C/16294T	0	11
16126C/16069T	0	10
16224C/16311C	0	12
16189C/16356C	0	6
16192T/16270 T	0	8
CRS[a]	0	26
Unassigned	14	24

[a] Cambridge reference sequence and those stemming directly from it.

East, central Asian populations have higher mean pairwise differences than are seen in European and western Asian populations, which would lend support to this hypothesis. However, an expansion from central to eastern Asia does not seem compatible with the higher mean pairwise differences and nucleotide polymorphism. Moreover, the simultaneous colonization of Europe and eastern Asia from central Asia might have implied overlapping mtDNA sequence pools between the two colonized regions, which is not the case.

Hypothesis B. Central Asian mtDNA Sequences: The Result of Admixture between European and Eastern Asian Sections

Intermediate nucleotide diversity, pairwise-difference means, and genetic distances, as well as a slightly elevated sequence diversity, are compatible with the central Asian mtDNA pool being an admixture of eastern Asian and European lineages. As shown in the Results section, Europe and eastern Asia share very few mtDNA sequences, whereas central Asia shares sequences with both groups of populations; the sequences found in our sample as well as in eastern Asia are different than the sequences shared by Europeans and central Asians. Moreover, of the two sequences shared by all four central Asian samples, one (identical to the Cambridge reference sequence) is the most frequent in Europe, whereas the other (bearing 16223T and 16362C) is the most frequent in eastern Asia. We found control-region sequences in central Asia that bear the motifs that are often associated with Asian RFLP haplotypes A–D (Torroni et al. 1993) and with European haplogroups H, J, K, T, V, and W (Torroni et al. 1996, 1998). However, it is extremely difficult to associate control-region sequences with the Asian superhaplogroup M (A. Torroni, personal communication). If central Asian mtDNA sequences are assumed to have a mixed origin, then the proportion of European sequences can be estimated as being 35%–48%, depending on the method, with the lower limit presumably being the more accurate estimate.

The dynamics of the process that generated the mixture of eastern Asian and European mtDNA sequences in central Asia is less clear. The same end result could be achieved through gene flow from the East and the West along the main trade routes during many generations, by migrations of whole groups during a shorter time span, or by a combination of both processes. Genetic distances and their representation by principal-coordinate analysis point to Mongols and/or Chinese as the possible source of eastern Asian sequences in central Asia. The Silk Road crossed the region, and it could have channeled migration along the east-west axis of Eurasia. However, we cannot exclude that admixture took place either before the establishment of the Silk Road or after its demise. Both the fact that an extinct Indo-European language, Tocharian, was spoken in the area during the latter half of the 1st millennium A.D. (Ruhlen 1991, pp. 35–36) and the recent discovery of mummified bodies with European facial traits point to the presence of Western peoples in central Asia. The analysis of mtDNA in those mummies could add additional support to the admixture hypothesis. Thus, although it is not possible to pinpoint the process that generated the central Asian peoples, it is clear that the different gene pools that merged in their formation had already diverged in the outer reaches of the Eurasian continent.

A further issue that can be investigated with the results that we have presented is the relation between the Turks and central Asian populations. Both the Turks and the

Table 3

Frequency of Sequences Identified as Eastern Asian or European, in Four Central Asian Populations

	No. (%) of Sequences		
	Eastern Asian	European	Unassigned
Kazakh	31 (56.4%)	22 (40.0%)	2 (3.6%)
Talas Kirghiz	35 (72.9%)	12 (25.0%)	1 (2.1%)
Sary-Tash Kirghiz	28 (59.6%)	15 (31.9%)	4 (8.5%)
Uighur	30 (54.5%)	19 (34.5%)	6 (10.9%)
Total	124 (60.5%)	68 (33.2%)	13 (6.3%)

Note.—The assignment of sequences was based, first, on the presence of the defining positions for eastern Asian and European groups (see table 2) and, second, on direct comparison with an extended set of eastern Asian and European mtDNA sequences.

populations that we have studied speak languages belonging to the Turkic branch of the Altaic family. During the 11th century A.D., Turkic nomads (such as the Seljuqs and the Ottomans, among others) occupied the grassland in the interior of Asia Minor, imposing their language and replacing Anatolian, an extinct branch of the Indo-European family (Ruhlen 1991, pp. 35–36), by an *élite dominance* process (Renfrew 1987, pp. 131–133). Whereas the historical and cultural consequences of the Turkic invasion of Anatolia were profound, the genetic contribution of the Turkic peoples to the modern Turkish population seems less significant. Previous studies (Calafell et al. 1996; Comas et al. 1996) have shown that the mtDNA pool found in Turkey can be interpreted as the result of upper Paleolithic and/or Neolithic expansions from the Middle East to Europe, with a small contribution by Asian sequences. The present results show that those sequences were found in the Turkic central Asian peoples, whose ancestors may have brought the Asian mtDNA sequences to Anatolia. Nevertheless, it should be stressed that, in the study of mtDNA sequences, only the female lineages are taken into account, whereas processes such as invasions by nomadic peoples might have been carried out basically by male warriors. Therefore, although, in the present study, we found evidence for a small influx of female lineages from central Asia to Turkey, a more complete picture of the history of Turkic populations could emerge if nuclear and Y-chromosome markers were analyzed.

Acknowledgments

This research was supported by Dirección General de Investigación Científico Técnica (Spain) grants PB92-0722 and PB95-0267-C02-01, by Human Capital and Mobility contracts (networks ERC-CRHX-CT920032 and ERB-CHRX-CT920090) to J.B., by Direcció General de Recerca, Generalitat de Catalunya grants 1995SGR00205 and 1996SGR00041, and by the Institut d'Estudis Catalans. This work was also made possible thanks to Spanish Ministry of Education and Science fellowship FI/93-1151 to D.C., FP93 to A.P.-L., and AP96 to R.M.-A.; Comissionat per a Universitats i Recerca, Generalitat de Catalunya fellowship FI/96-1153 to E.B. and postdoctoral fellowship to F.C.; and the University of Barcelona fellowships to E.M. and J.C. We are also thankful for the help and technical assistance offered by the staff of the Unitat de Seqüenciació, Servei Científico-Tècnic, Universitat de Barcelona. The Laboratory of Anthropology of the Academy of Sciences of Kazakhstan (Prof. O. Ismagulov and Dr. A Ismagulova) is gratefully acknowledged for help with the fieldwork. The CAHAP research project was partially supported by MURST grants to F.F. and D.P. and by private grants from AGIP SpA (Milano). We also thank the people of Sary-Tash, Talas, Keghen, and Penjim for their hospitality and cooperation.

Appendix

mtDNA Control-Region Segment I Sequences in Central Asian Populations

In the Appendix figure (displayed on the following two pages), dots (.) denote identity with the reference sequence (AND [Anderson et al. 1981]), and numbers at the end of each sequence indicate the absolute frequencies of each sequence in the Kazakhs (KAZ), the highland Kirghiz (KIR), the lowland Kirghiz (KIT), and the Uighurs (UIG). Each sequence either has been considered to be of eastern Asian origin (denoted by an "A" at the end of the line) or European origin (denoted by an "E" at the end of the line) or has been left unassigned (denoted by an "X" at the end of the line).

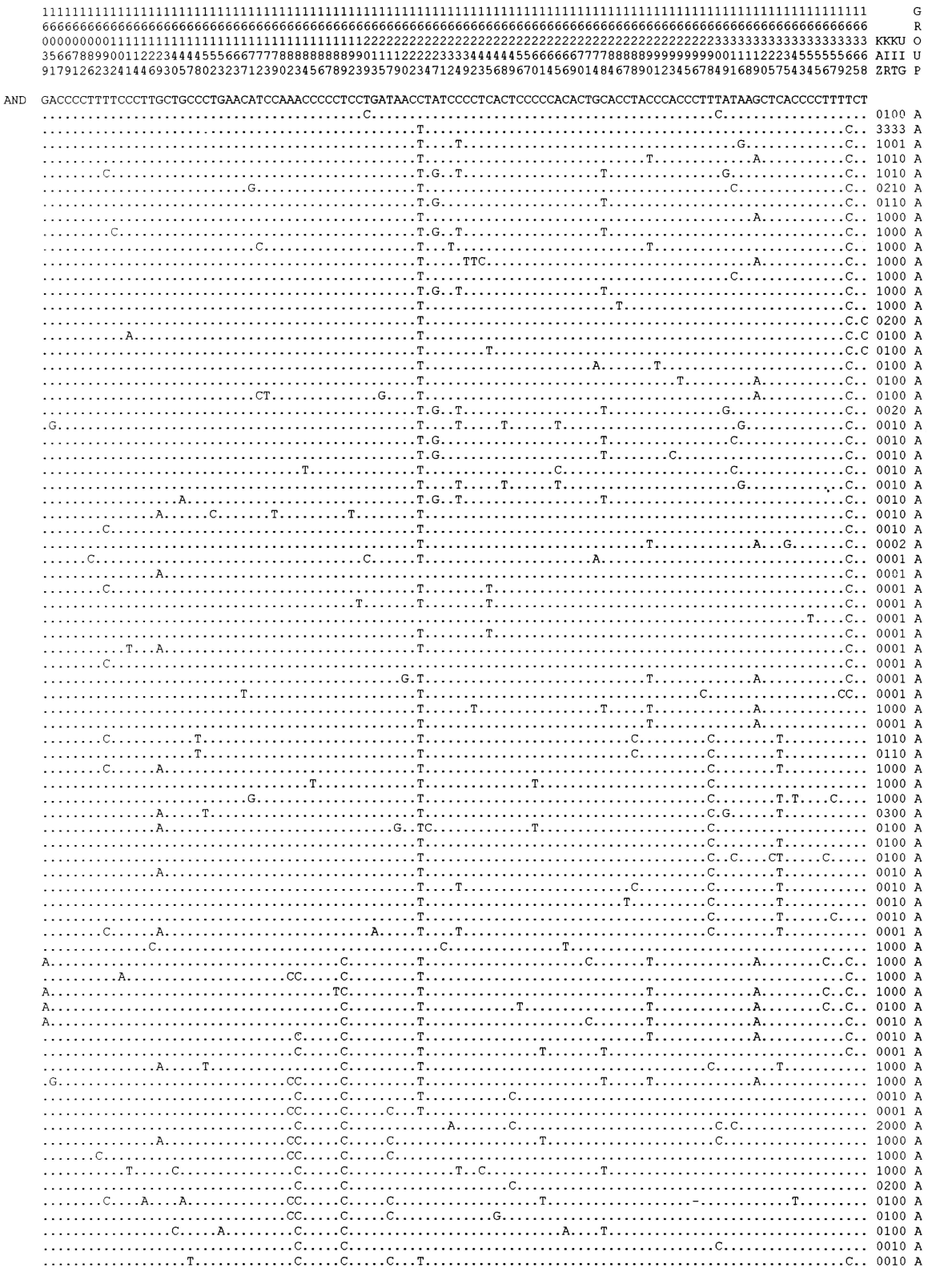

Figure A1

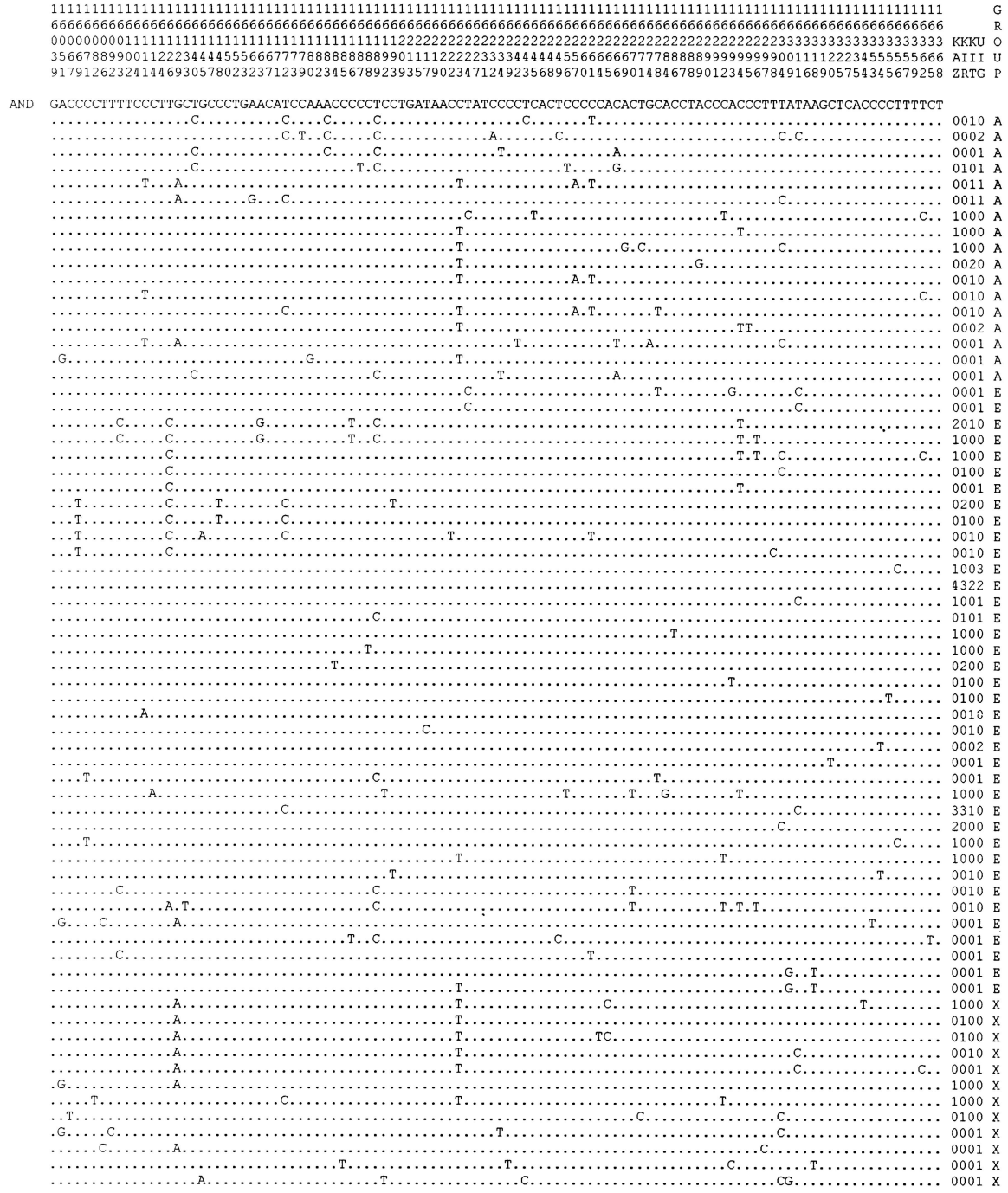

Figure A1 Continued

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–465

Anthony DW, Brown DR (1991) The origins of horseback riding. Antiquity 65:22–38

Anthony DW, Vinogradov (1995) Birth of the chariot. Archaeology 28(2): 36–41

Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13:494–504

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen K-H, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics 130:139–152

Bandelt H-J, Forster P (1997) The myth of bumpy hunter-gatherer mismatch distributions. Am J Hum Genet 61: 980–983

Bandelt H-J, Forster P, Sykes B, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Batista O, Kolman CJ, Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panama. Hum Mol Genet 4:921–929

Battistini N, Facchini F, Bedogni G, Severi S, Fiori G, Pettener D (1995) The prediction of extracellular and total body water from bioelectric impedance in a non-Caucasian population from central Asia. Ann Hum Biol 22:315–320

Bedogni G, Battistini N, Severi S, Facchini F, Pettener D, Fiori G (1997) Body water distribution in Kirghiz highlanders versus lowlanders. Ann Hum Biol 24:533–538

Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995) Human mitochondrial DNA variation and the origin of the Basques. Ann Hum Genet 59:63–81

Bowles GT (1977) The peoples of Asia. Weidenfeld & Nicolson, London

Cabot EL (1988) ESEE: the eyeball sequence editor, version 1.06. University of British Columbia, Burnaby

Calafell F, Underhill PA, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann Hum Genet 60:35–49

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton

Comas D, Calafell F, Mateu E, Pérez-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, Pérez-Lezaun A, Bosch E, Bertranpetit J (1997) Mitochondrial DNA variation and the origin of the Europeans. Hum Genet 99:443–449

Côrte-Real HBSM, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, et al (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331–350

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Efron B (1982) The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87–112

Facchini F, Pettener D, Rimondi A, Sichimbaeva K, Riva P, Salvi P, Pretolani E, et al (1997) Taste sensitivity to PTC and thyroid function (FT4 and TSH) in high- and low-altitude Kirgiz populations in the Pamir. Hum Biol 69:97–106

Facchini F, Toselli S, Ismagulov O, Fiori G, Ismagulova A, Pettener D (1998) Body composition in central Asian populations: I. Fat patterning variation in the Kazakhs of the Tien Shan mountains (2,100 m) and the Uighurs of the Semericia. Am J Hum Biol 10:241–247

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 35:785–791

——— Felsenstein J (1989) PHYLIP: phylogeny inference package, version 3.2. Cladistics 5:164–166

Francalacci P, Bertranpetit J, Calafell F, Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100:443–460

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325–338

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345

Handt O, Meyer S, von Haeseler A (1998) Compilation of human mtDNA control region sequences. Nucleic Acids Res 26:126–129

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, et al (1994) Molecular genetic analyses of the Tyrolean Ice Man. Science 264:1775–1778

Harpending H, Sherry ST, Rogers AR, Stoneking M (1993) The genetic structure of ancient human populations. Curr Anthropol 34:483–496

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, et al (1996) mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590

Howell N, Kubacha I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501–509

Howell N, Mackey D (1997) Reply to Macaulay et al. Am J Hum Genet 61:986–990

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngobe Amerinds of Panama. Genetics 140:275–283

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

Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E, Sykes B, Bandelt H-J (1997) mtDNA mutation rates: no need to panic. Am J Hum Genet 61:983–986

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. Genetics 97:145–163

Pääbo S (1996) Mutational hot spots in the mitochondrial microcosm. Am J Hum Genet 59:493–496

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Allliston-Greiner R, Wilson MR, Berry DL, et al (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368

Pettener D, Facchini F, Luiselli D, Toselli S, Rimondi A, Ismagulova A, Sichimbaeva K, et al (1997) Physiological adaptability, thyroid function, body composition and genetic variability in central Asia high altitude populations. Acta Andina 6:217–226

Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 106:85–90

Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60: 321–330

Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Pääbo S (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Sayler 375:837–840

Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. Theor Popul Biol 21:24–43

Renfrew C (1987) Archaeology and language: the puzzle of Indoeuropean origins. Jonathan Cape, London

Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, et al (1996) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569

Ruhlen M (1991) A guide to the world's languages. Stanford University Press, Stanford

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, et al (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Res 5:42–52

Santos M, Ward RH, Barrantes R (1994) mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. Hum Biol 66:963–977

Schneider S, Kueffer JM, Roessli D, Excoffier L (1996) Arlequin: a software environment for the analysis of population genetics data, version 1.0. Genetics and Biometry Lab, University of Geneva, Geneva

Sellier J, Sellier A (1993) Atlas des peuples d'Orient. Editions de la Découverte, Paris

Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. Am J Hum Genet 53:549–562

Stenico M, Nigro L, Bertorelle G, Calafell F, Capitanio M, Corrain C, Barbujani G (1996) High mitochondrial sequence diversity in linguistic isolates of the Alps. Am J Hum Genet 59:1363–1375

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595

Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, et al (1998) mtDNA analysis reveal a major late paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62: 1137–1152

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, et al (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. Am J Phys Anthropol 93:189–199

Torroni A, Sukernik, RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MF, Comuzzie AG, et al (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. 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

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720–8724

Ward RH, Redd A, Valencia D, Frazier BL, Pääbo S (1993) Genetic and linguistic differentiation in the Americas. Proc Natl Acad Sci USA 90:10663–10667

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV (1996) Mitochondrial DNA polymorphism in three Brazilian Indian tribes. Am J Hum Biol 8:317–323

Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S (1996) mtDNA sequence diversity in Africa. Am J Hum Genet 59:437–444

Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704