

ARTICLE

Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages

David Comas^{*,1,4}, Stéphanie Plaza^{1,4}, R. Spencer Wells², Nadira Yuldaseva^{2,3}, Oscar Lao¹, Francesc Calafell¹ and Jaume Bertranpetit¹

¹Unitat de Biologia Evolutiva, Departament de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona 08003, Spain; ²Wellcome Trust Center for Human Genetics, University of Oxford, Headington, UK; ³Institute of Immunology, Academy of Sciences, Tashkent, Uzbekistan

Mitochondrial DNA (mtDNA) lineages of 232 individuals from 12 Central Asian populations were sequenced for both control region hypervariable segments, and additional informative sites in the coding region were also determined. Most of the mtDNA lineages belong to branches of the haplogroups with an eastern Eurasian (A, B, C, D, F, G, Y, and M haplogroups) or a western Eurasian (HV, JT, UK, I, W, and N haplogroups) origin, with a small fraction of Indian M lineages. This suggests that the extant genetic variation found in Central Asia is the result of admixture of already differentiated populations from eastern and western Eurasia. Nonetheless, two groups of lineages, D4c and G2a, seem to have expanded from Central Asia and might have their Y-chromosome counterpart in lineages belonging to haplotype P(xR1a). The present results suggest that the mtDNA found out of Africa might be the result of a maturation phase, presumably in the Middle East or eastern Africa, that led to haplogroups M and N, and subsequently expanded into Eurasia, yielding a geographically structured group of external branches of these two haplogroups in western and eastern Eurasia, Central Asia being a contact zone between two differentiated groups of peoples.

European Journal of Human Genetics advance online publication, 11 February 2004; doi:10.1038/sj.ejhg.5201160

Keywords: mitochondrial DNA; genetic admixture; haplogroup; Central Asia

Introduction

Central Asia is a vast territory that has been crucial in human history due to its strategic location. Situated eastwards of the Caspian Sea, limited by the Hindu Kush and Altai mountain ranges to the east and by the great Asian Steppes to the north, this territory has been a complex assembly of peoples, cultures, and habitats.

The area has been occupied since Lower Paleolithic times, and there is evidence of Neanderthal skeletal

material in Teshik-Tash,¹ Uzbekistan. Nonetheless, the later expansion of Upper Paleolithic remains is far less clear.² Classical Greek and Chinese historic records cite the Scythians and Sarmatians, Indo-European-speaking people described as having European morphological traits, as the first inhabitants occupying the region. These historic citations raise the questions of the origin of the ancestors of the modern settlers across the region, and of the limits of western peoples in Asia. Several facts point to the presence of western peoples far east in Asia, such as an extinct Indo-European language (Tocharian) spoken during the latter half of the first millennium in Chinese Turkestan, the presence of mummified bodies with European facial traits in the Xinjiang region, the description of west Eurasian mitochondrial DNA lineages in Central Asia,³ and the suggested European affiliation of mitochondrial

*Correspondence: Dr D Comas, Universitat Pompeu Fabra Unitat de Biologia Evolutiva, Doctor Aiguader 80 Barcelona, Catalonia 08003, Spain. Tel: +34 93 5422844; Fax: +34 93 5422802; E-mail: david.comas@upf.edu

⁴These authors contributed equally to the present study. Received 28 July 2003; revised 12 December 2003; accepted 17 December 2003

DNA sequences from ancient bones in an Eastern Chinese site.⁴ Besides Scythians and Sarmatians, other peoples left their influence in the area: Greeks, Chinese, Turkic tribes such as the Huns, and the Avars, Arabs, and others.

Physical anthropology has roughly defined Central Asian populations as presenting an admixture of eastern and western anthropometric traits.⁵ There are few genetic data about the human populations settled in the region. Classical genetic data⁶ have demonstrated an intermediate position of Central Asians between the Middle East and East Asia. As a general rule, the people inhabiting the area are the result of admixture between differentiated populations, which has produced a high genetic diversity.^{3,7-9} Nonetheless, recent data of Y-chromosome lineages in Central Asia¹⁰ have shown that genetic diversity is heterogeneous in the region, with some high-diversity populations contrasting with much reduced levels in others. This pattern has been interpreted as the occurrence of several bottlenecks or founder events in the area.

Mitochondrial DNA (mtDNA) lineages have been used to unravel past demographic scenarios due to their particular properties. Previous mtDNA analyses in Central Asia based on the sequence of the first hypervariable segment of the control region³ have shown that the mtDNA pool of three populations in Central Asia (the Kazakh, the Kirghiz, and the Uighur) is the result of admixture from east and west Eurasia. Although mtDNA control region sequences allowed the general distinction between the Eastern and Western sources, it did not allow full resolution into haplogroups and of the phylogeographic perspective. The knowledge provided by complete mtDNA sequences¹¹⁻¹⁵ and the refined definition of haplogroups both in West Eurasia¹⁶ and in East Asia^{15,17} provides a fine-grained phylogeography of the mtDNA lineage distribution, which might allow us to determine which mtDNA markers should

be determined to analyze the diversity of the present Central Asian samples.

The analysis of extant central Asians allows us to test several scenarios concerning the spread of western peoples in Asia and their interaction with eastern peoples. In this sense, we have analyzed 12 populations from all the major linguistic groups in the area, and have typed both hypervariable segments of the control region as well as some key SNPs in order to achieve a much finer phylogeographic resolution. This will allow a more complete description of the mtDNA diversity in Central Asia, and its interpretation in relation to human origins and dispersals into and out of Central Asia.

Material and methods

A total of 232 individuals from 12 different population groups were analysed: 20 Bukharan Arabs, 20 Crimean Tatars, 20 Iranians, 16 Dungans, 20 Karakalpaks, 20 Kazaks, 20 Khorezmian Uzbeks, 20 Kyrgyz, 20 Tajiks, 20 Turkmen, 16 Uighurs, and 20 Uzbeks. Samples were collected in Uzbekistan and Kyrgyzstan, with informed consent; information about the origin of maternal ancestors was recorded in order to localize samples geographically, and their locations are shown in Figure 1.

DNA was extracted from blood samples using standard methods. Both mtDNA hypervariable regions (HVRI and HVRII) were amplified using primers L15996 and H408,¹⁸ and the amplification products were subsequently purified with the GenClean (BIO101) kit. The sequence reaction was performed for each strand, using primers L15996 and H16401 for the HVRI, and L29 and H408 for the HVRII,¹⁸ with the ABI PRISM dRhodamine Terminator Cycle Sequencing kit (Applied Biosystems) according to the supplier's recommendations. Sequences from positions

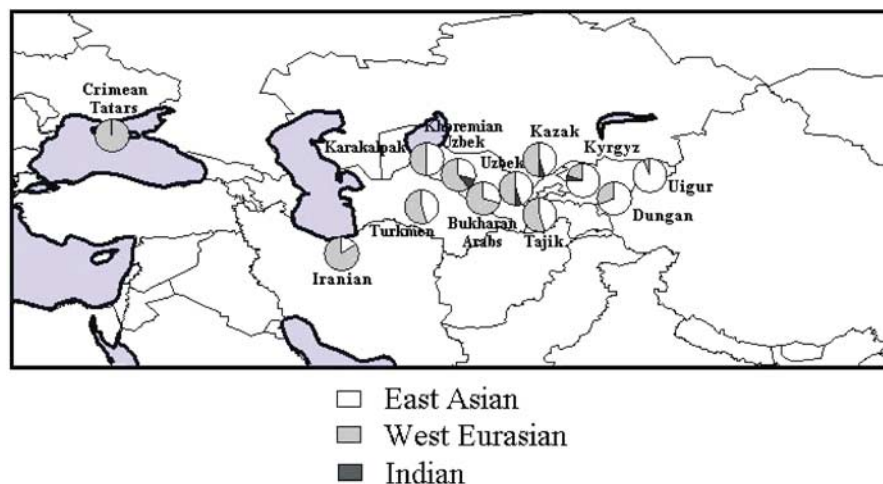


Figure 1 Geographic location of samples analyzed in the present study. Frequencies of East Asian, West Eurasian, and Indian lineages are shown in white, pale gray, and dark gray, respectively.

16024–16391 and 63–322, respectively,^{19,20} were obtained.

The 9-bp tandem repeat (CCCCCTCTA) of the COII/tRNA^{Lys} intergenic region was amplified using primers L8196 (5'-ACAGTTTCATGCCCATGGTC-3', labeled at 5' with JOE) and H8297 (5'-ATGCTAAGTTAGCCTTACAG-3'). The cycling conditions were as follows: 94°C for 2 min; followed by 30 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min; and a final elongation step of 72°C for 5 min. The product was run in an ABI PRISM377 and GeneScan analysis software was used to measure the fragment sizes.

Three positions in the mtDNA coding region^{19,20} (10 400, 12 308, and 12 705) were also determined by using the SNaPshot™ ddNTP Primer Extension Kit (Applied Biosystems), which consists of a single-base primer extension which uses labeled ddNTPs to interrogate SNPs. The mtDNA region containing the three SNPs was amplified using primers L10373 (5'-CCCTAAGTCTGGCCTATGAG-3') and H12744 (5'-CGATGAACAGTTGGAATAGG-3'), with the following cycling conditions: 94°C for 5 min; 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s; and a final elongation step of 72°C for 5 min. The 2410-bp amplification products were purified using the QIAquick™ PCR Purification Kit (QIAGEN). The single-base primer extension was performed following supplier's recommendations using oligonucleotides H10400X (5'-TGTTAAACTATA-TACCAATTC-3'), L12308X (5'-CAGCTATCCATTGGTCT-TAGGCCCAA-3'), and L12705X (5'-AACATTAATCAGTT-CTTCAAATATCTACTCAT-3') in the same reaction. Unincorporated-labeled ddNTPs were removed by adding 1 U of CIP to the primer extension products for 1 h at 37°C, followed by an incubation of 15 min at 72°C to inactivate the enzyme. Products were run in an ABI PRISM377 and

GeneScan analysis software was used to measure fragment sizes.

Each mtDNA molecule was assigned to one haplogroup according to the following strategy. First, the combination of the three SNPs in the coding region was taken into account in order to classify the mtDNA molecules in one of the four major groups determined in the present work: R, U, M, or other (namely, L or N). Subsequently, the information yielded by the control region sequence was added in order to refine the classification into haplogroups^{15–17} (see Figure 2). Nonetheless, after this assignment strategy, some individuals were difficult to be classified as N or L3. For this reason, variation at position 10 873, distinguishing haplogroup N from L3, was also tested using the single-base primer extension approach with oligonucleotide L10873X (5'-TTTTTTTTTCCA-CAGCCTAATTATTAGCATCATCCC-3').

In order to compare the present results with other populations, HVRI data from several European, Middle Eastern, Indian, Central Asian, and East Asian populations were taken from the literature: Kazaks,³ Kyrgyz,³ Uighurs,³ Altaics,²¹ Mongolians,^{22,23} Daur,²³ Oroqen,²³ Turks,^{24–26} Han Chinese,^{17,27} Han Taiwanese,²⁸ Ainu,²⁸ Koreans,^{23,28,29} Japanese,^{28,30} Europeans,³¹ Middle Easterns,³¹ Caucasus populations,^{32,33} Thai,³⁴ Indians,³⁵ Russians,^{36,37} Ukrainians,³⁷ and Siberians.^{23,38,39}

The networks relating HVRI sequences within some of the haplogroups described were constructed by using a reduced-median algorithm⁴⁰ as implemented in the Network 3.0 program. The dating method employed⁴¹ is based on the average number of mutations accumulated from an ancestral sequence as a linear function of time and mutation rate. This method was also performed with the Network 3.0 program.

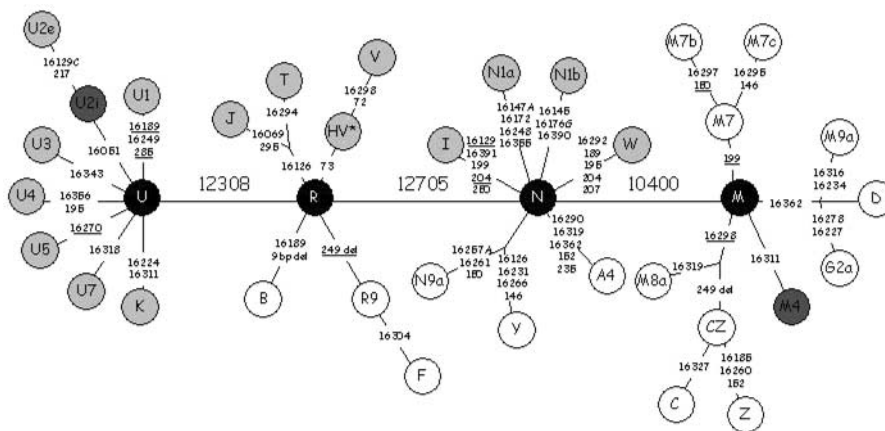


Figure 2 Phylogenetic reconstruction and geographic distribution of the haplogroups found in Central Asia. Numbers along the links indicate substitutions (transversions are indicated by the substituted nucleotide after the number), underlined numbers indicate recurrent events. East Asian, West Eurasian, and Indian lineages are shown in white, pale gray, and dark gray, respectively.

Program Admix 2.0⁴² was used to calculate the admixture proportions of the present samples based on the frequency of the haplogroups. As putative parental populations, we used four data sets that consisted of 258 Eastern Europeans³¹ (Bulgarians, Romanians, and Russians), 316 Middle Easterns³¹ (Bedouins, Syrians, and Turks), 190

Northern Indians, and Pakistanis³⁵ (regions of Uttar Pradesh, Rajasthan, Punjab, Kashmir, Haryana, and Pakistan), and 263 East Asians²⁷ (Han Chinese).

In order to detect the possible genetic structure among populations, an analysis of the molecular variance (AMOVA)⁴³ was performed using the Arlequin package.⁴⁴

Table 1 Haplogroup frequencies in the samples analyzed.

	CT	IR	TU	KR	KU	AR	UZ	TD	KZ	KG	DU	UI	Total
<i>R</i>													
HV*	4	3	6	5	5	7	3	3	4		1	1	42
V						1				1			2
J	2	2	1	1	1	2	1	1	1				12
T	6	1	1	1	1	1	1	1	1		3		17
B			1		2	1	1	1			1		7
F			1	1			1			3	1	4	11
R9				1			1						2
R*		2	1	1			3		1		1		9
<i>U</i>													
K	2	2			1		1	1					7
U1	1	1		2		2		1	1				8
U2e						1				1			2
U2i					2		1			1			4
U3	2												2
U4	1				2		1						4
U5		1			1			1	2				5
U7								1					1
<i>M</i>													
CZ				1									1
C		2	1			4	2		1	6	1		17
Z				1					1				2
D (D4c)		1	6 (5)	4	2		1	3 (2)	2 (1)	4	1	7(2)	31(10)
G2a				1			2	1	4				8
M4									1				1
M7b										1	2	1	4
M7c				1								1	2
M8							1			1		1	3
M9											1		1
<i>Other</i>													
I	1	1			1					1			4
W								2		1			3
N1a			1										1
N1b	1	3											4
N9a											2		2
A4					2			3			2	1	8
Y						1		1	1				3
N*		1	1										2
East	—	3	9	10	6	6	9	9	9	15	11	15	102
M	—	0.15	0.45	0.50	0.28	0.30	0.44	0.45	0.44	0.74	0.69	0.94	0.48 ^a
West	20	17	11	10	12	14	10	11	10	4	5	1	125
M	1	0.85	0.55	0.50	0.54	0.70	0.45	0.55	0.45	0.17	0.31	0.06	0.48 ^a
India	—	—	—	—	2	—	1	—	1	1	—	—	5
M	—	—	—	—	0.17	—	0.09	—	0.09	0.09	—	—	0.04 ^a
Total	20	20	20	20	20	20	20	20	20	20	16	16	232

In D haplogroup, in parenthesis, individuals belonging to D4c haplogroup. CT: Crimean Tatars, IR: Iranian, TU: Turkmen, KR: Karakalpak, KU: Khoreman Uzbek, AR: Bukharan Arabs, UZ: Uzbek, TD: Tajik, KZ: Kazak, KG: Kyrgyz, DU: Dungan, and UI: Uighur, m: estimated contribution of each region (East Asia, West Eurasia, and India) to each Central Asian population, taking into account that the Indian mtDNA pool contains also East Asian and West Eurasian sequences, a Total contribution without taking into account Crimean Tatars.

Results

Phylogeographic structure

A total of 232 individuals have been analyzed for the HVRI and HVRII, for the presence of the 9-bp tandem repeat of the COII/tRNA^{Lys} intergenic region, and several SNPs in the mtDNA coding region. Individual data are available in the following web site (<http://www.upf.es/cexs/bioevo/index.html>).

Haplogroup frequencies by population are shown in Table 1. In all, 11 sequences were difficult to assign to a specific haplogroup and were named after the first major classification yielded by the coding SNPs (all belong either to R* or N*). The haplogroups found and the positions that define them are shown in Figure 2.

Within the present samples, no African lineages were found. No sub-Saharan L (L1, L2, and L3) lineages^{45,46} were present in Central Asian samples. Other haplogroups of African origin, such as U6 from North Africa,⁴⁷ or M1 from East Africa,⁴⁸ are not found in the present sample set.

Within major group R, mtDNA molecules analyzed belong either to West Eurasian haplogroups (H, V, J, and T) or to East Asian haplogroups (B, R9, and F). Within this group of lineages, the West Eurasian haplogroup HV* (including pre-HV, HV, and H) is the most numerous, and it is present in all the analyzed populations except the Kyrgyz. Two individuals belong to haplogroup V, which is likely to be of Western European origin.⁴⁹ Nevertheless, the range of haplogroup V extends far beyond Europe, into Northern Africa⁵⁰ and as far East as Central Asia.

MtDNA molecules belonging to major group U have their origin in West Eurasia and they have been found in most Central Asian populations. Nevertheless, Kivisild *et al*³⁵ distinguished two groups of lineages within haplogroup U2: West European U2e and Indian U2i. Within the present sample set, we have found both the U2 groups.

In continental Asia, lineages belonging to major group M have an Indian (M2, M3, M4, M5, and M6)⁵¹ or an East Asian (C, D, E, Z, M7, M8, M9, M10, and M11)^{15,17} origin. Only one M Indian lineage (belonging to the M4 haplogroup) has been found in the sample set, whereas the rest of M lineages have an East Asian origin. Haplogroup D is the most frequent haplogroup within this major group, followed by C lineages. Some mtDNA molecules belonging to E and G root lineages might have been classified as D since they are not distinguishable by control region sequence substitutions; this is not a major bias as all of them are of East Asian distribution.

Within D, a non-negligible fraction of sequences carry a transition at position 16245. This group may be a clear subclade within D, which, pending further coding-region characterization, we suggest to call D4c. D4c is highly frequent and diverse in Central Asia (25% in Turkmen, 10% in Tajik, 7% in Uighur, 2.7% in Kazak, and 0.9% in Kyrgyz) (present data and Comas *et al*³), and it is found at low

frequencies, in Turks (2.1%), Daur (8.9%, only two sequences), Mongolians (0.7%), southern Siberians (0.7%), Han Chinese (0.6%), and Koreans (0.5%). This group of sequences is absent in other East Asian, Indian, and Middle Eastern samples. The structure of the variation of these sequences is shown as a network in Figure 3, from which an age of 25 000 (SE 9600) years can be estimated.

All G lineages found in the present samples belong to the G2a group; thus, no G1 or G3 lineages were found. In fact, the presence of G2a lineages seems to be also restricted to Central Asia. This haplogroup characterized by the motif 16 223T, 16 227G, 16 278T, and 16 362C, has been found in Kazaks (9.3%), Kyrgyz (7.0%), Karakalpak (5.0%), Tajik (5.0%), and Uzbek (5.0%) (present data and Comas *et al*³). It has also been found in neighboring populations at lower frequencies, such as Mongolians (1.3%), Mansi from Siberia (6.1%, only one sequence), southern Siberians (2.4%), Ainu (3.9%), Japanese (0.7%), Daur (4.4%, two sequences), Han Taiwanese (3.0%), Korean (1.9%), Han Chinese (2.2%), and the Caucasus (0.6%). The structure of the variation of haplogroup G2a is shown in Figure 4, from which an age of 29 500 (SE 7000) years can be estimated.

Other haplogroups found in Central Asia are A, Y, and N9a, which have an East Asian origin, whereas haplogroups W, I, N1a, and N1b have been described in West Eurasian populations.

Admixture analysis

The presence in Central Asia of a high proportion of sequences originating elsewhere suggests that these populations have experienced intense gene flow. In order to quantify the apportionment of admixture in Central Asian samples, two different approaches were followed: a phylogeographic approach and an admixture approach based on haplotype frequencies. Crimean Tatars were excluded from the admixture analysis since their geographic position corresponds more to Europe rather than Central Asia, and their mtDNA pool is completely of West Eurasian origin.

Taking into account the phylogeography of the haplogroups described for West Eurasia¹⁶ and East Asia,^{15,17} these can be divided into three groups depending on their origins: West Eurasian, East Asian, and Indian (Table 1 and Figure 1). Whereas West Eurasian and East Asian populations contain almost exclusively locally originated mtDNA haplogroups, this is not the case for India. Then, admixture from India would also contribute West and East Eurasian sequences to Central Asia. Thus, estimated admixture proportions have been corrected with the frequencies of haplogroups of Indian (58.4%), West Eurasian (32.6%), and East Asian (8.9%) origins in a sample from India and Pakistan.³⁵ Standard deviations were estimated by sampling with replacement 100 000 times in samples having the same sizes and haplogroup frequencies as those in Central Asia and India, and computing each time the admixture proportion estimates. Considering all the

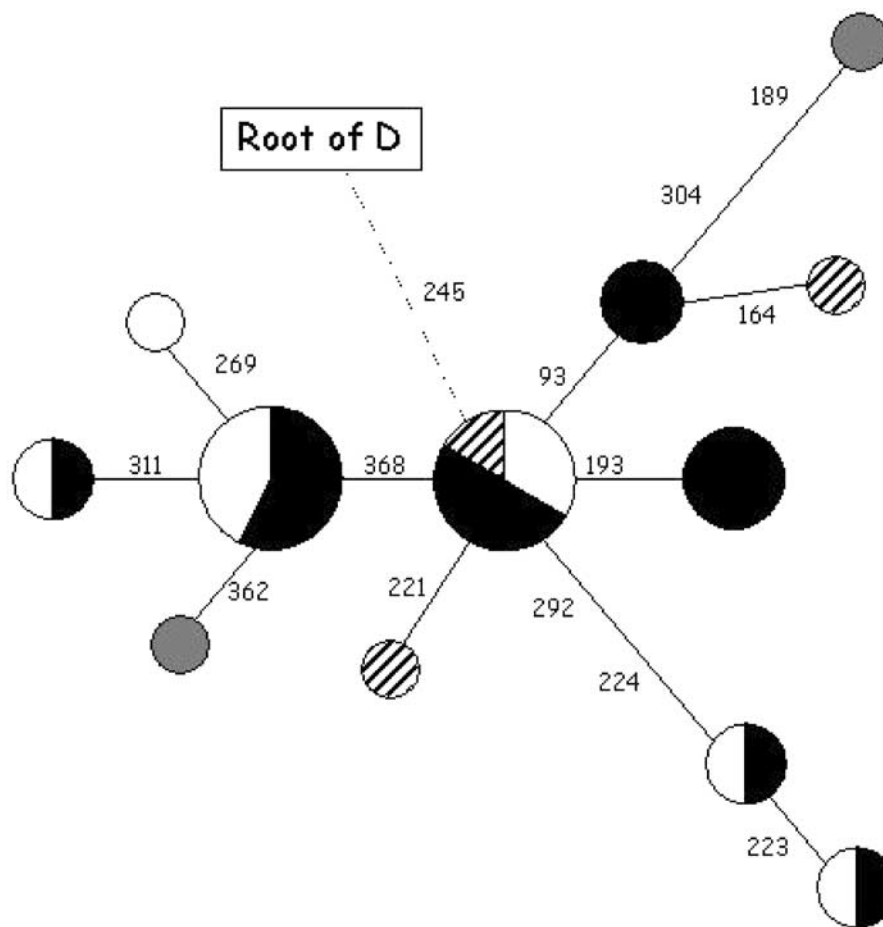


Figure 3 Phylogenetic network of a section of haplogroup D sequences (D4c). The size of the circles is proportional to the number of sequences. Central Asian samples are represented in black, East Asians in white, Turks in gray, and Siberians in striped gray. Mutated sites (minus 16 000) are indicated along the lines.

individuals as belonging to a single hybrid population, the estimated admixture proportions are 0.48 ± 0.04 West Eurasian, 0.48 ± 0.04 East Eurasian, and 0.04 ± 0.02 Indian. Given the sample sizes for individual populations, their admixture proportions (Table 1) carry large standard errors and are not discussed separately.

An admixture approach⁴² was performed using the method implemented in Admix 2.0 program, considering four putative parental populations. The apportionment for the whole sample set was 0.11 ± 0.24 European, 0.40 ± 0.25 Middle Eastern (which adds up to 51% for West Eurasia), 0.45 ± 0.05 East Asian, and 0.04 ± 0.04 Indian. Although this approach allowed us to use a larger number of parental populations, the standard deviation after 10 000 iterations is extremely high for some of the estimates.

Genetic structure of Central Asia

The genetic structure of Central Asian populations was investigated through AMOVA. When the 12 samples were considered as a single group, only 2.34% ($P < 0.0001$) of the

genetic variance was attributed to differences among populations. When samples were grouped according to language families (Afro-Asiatic, Altaic, Indo-European, and Sino-Tibetan), the fraction of the genetic variance found among groups was not significantly different from 0 ($P = 0.817$), whereas differences found among populations within language groups were statistically significant (2.9%, $P < 0.0005$), showing that the genetic variation found in the mtDNA was not structured according to language affiliation.

Discussion

The mtDNA genetic landscape of Central Asia contains four main differentiated lineage groups according to their phylogeographic origin: (i) a group of lineages originating in West Eurasian and comprising almost half of the mtDNA sequences in Central Asia; (ii) East Asia lineages, making almost the other half of lineages, (iii) two putatively locally expanded haplogroups, of East Asian origin, D4c and G2a,

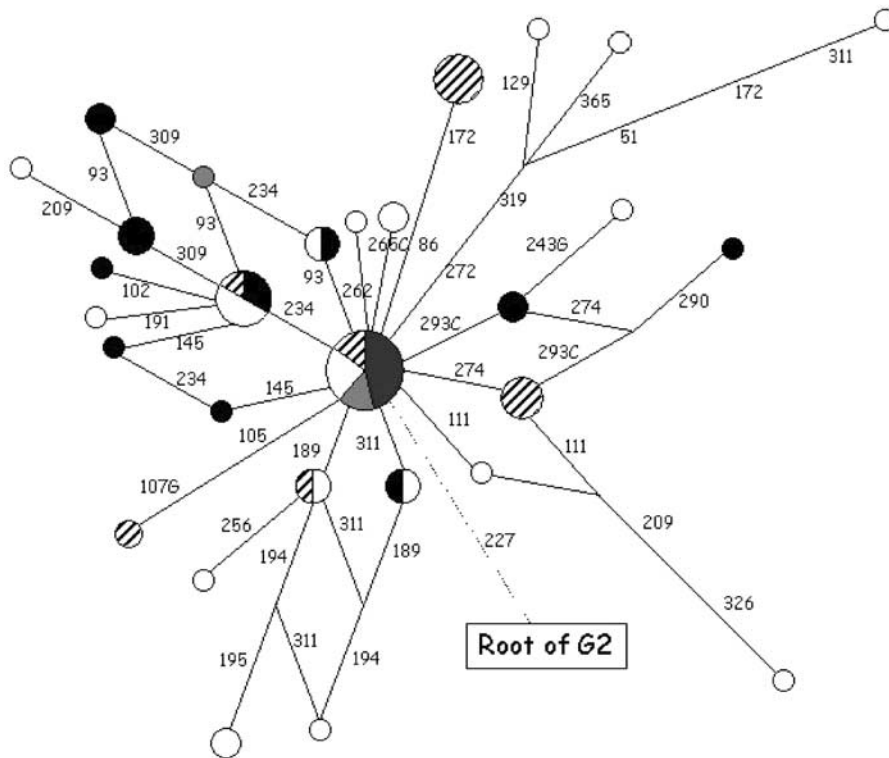


Figure 4 Phylogenetic network of haplogroup G2a. The size of the circles is proportional to the number of sequences. Central Asian samples and Mongolians are represented in back, East Asians in white, samples from the Caucasus in gray, and Siberians in stripped gray. Mutated sites (minus 16 000) are indicated along the lines.

accounting for a ~8% of the total sequences, and (iv) a tiny fraction of sequences of Indian origin.

We have detected some groups of sequences mainly restricted to this geographical area. This is the case of haplogroups G2a and D4c. The fact that these groups of lineages are localized in Central Asia at higher frequencies than in neighboring populations could be explained as a result of genetic drift during founder events that could have raised its frequency in this geographical area. Nevertheless, the high diversity found in Central Asia within both groups of sequences (Figures 3 and 4) supports an ancient origin of the founder mutations (around 30 000 and 25 000 years), an expansion of these lineages in Central Asia, and subsequent dispersal to neighboring populations. These ancient events represent ancient expansions originated in Central Asia and might have their Y-chromosome counterpart in lineages belonging to haplotype P(xR1a) that has a high frequency in Central Asia and is dated to $\approx 40\,000$ years.⁸ There is, thus, a fraction of the gene pool that can be considered Central Asian specific, which could reflect the remnants of the oldest peopling by modern humans.

Besides the specific cases of G2a and D4c lineages, no other lineages seem to have expanded in Central Asia, and the majority of lineages found have an Eastern or Western

origin, which are two mtDNA pools that do not overlap. This fact implies that both genetic pools were already differentiated when they met in Central Asia. Thus, the geographic distribution of mtDNA lineages in Europe and Asia is not compatible with a Central Asian origin of both mtDNA pools, in agreement with previous data.³

The presence of western sequences in Central Asia prompts the question of the eastern spread of western influence in Asia. The analyses performed of the ancient sites of Liangchun⁴ (2500 years old) and Yixi⁵² (2000 years old), eastern China, concluded that there was a drastic shift from a European-like population 2500 years ago, through an intermediate population 2000 years ago, to the present-day East Asian populations. Liangchun sequences are difficult to assign to haplogroups due to the short mtDNA sequence analyzed, and their ascription to the Western Eurasia gene pool has been challenged⁵³ up to the point that the latter authors do not interpret any Liangchun sequence as Western. On the other hand, most Yixi sequences belong to extant East Asian haplogroups such as D, C, or F, which suggests that the genetic composition of the 2000-year-old Yixi site presented no genetic traces of western influence. The genetic influence of western peoples across Asia is obvious in Central Asia, but there is no evidence of its presence in the easternmost regions

since no traces are found in extant or ancient East Asian populations. Even if Tocharian, an Indo-European language, was present in Eastern Asia, there is no evidence, from extant genetic variation in maternal lineages, of the Western Eurasia genetic contribution.

The presence of western and eastern sequences found in Central Asia leaves open questions about the mode and tempo of the generation of this admixture of lineages. Two scenarios could have produced this mtDNA pattern in Central Asia:

- (a) Western peoples inhabited Central Asia and were partially replaced by Eastern peoples, Central Asia being a hybrid zone.
- (b) Central Asia has been a 'contact zone' between two differentiated groups of peoples who originated in east and west Eurasia, respectively.

The revision of the ancient sequences from China⁵³ and the finding of specific Central Asian sequences clearly support the second. G2a and D4c haplogroups are 'twigs' (according to the terms devised by Kivisild *et al*¹⁷) belonging to the East Asian G and D 'limbs' of the M 'trunk'. The estimated ages of these haplogroups (around 30 000 and 25 000 years) point to the ancient presence of at least two different East Asian 'limbs' in Central Asia.

Kivisild *et al*¹⁷ showed considerable differences in the mtDNA lineages found in East Asia, A, C, D, G, Y, and Z being the haplogroups forming the pool of lineages in the northeast, whereas B and F were predominant in the southeast. Karafet *et al*,⁹ analyzing Y-chromosome markers, showed a closer genetic relationship between Central Asia and northeast Asia than with southeast Asia. Nevertheless, our mtDNA results show the presence of haplogroups represented in both northeast and southeast Asia, suggesting that the demographic scenario within Central Asia has been even more complex than previously stated.⁹

Contrary to the structure shown in Y-chromosome lineages in Central Asia, where 24% of the genetic variation could be attributed to differences between populations,¹⁰ mtDNA diversity is not structured, as shown by the AMOVA analysis. This discrepancy between the two uniparental genomic regions in Central Asia is in agreement with previous data in the region,⁷ and as a global trend in which higher female than male migration has been observed.⁵⁴

It is interesting to stress the lack of geographic structure of the basal branches of the non-African mtDNA (haplogroups M and N, called 'limbs'¹⁷), and a clear phylogeography in more external branches (haplogroups or sub-haplogroups; 'twigs'¹⁷) supports the existence of an ancestral population where the two main groups of lineages diverged. This could be related to the presence of a 'maturation phase', presumably in the Middle East or eastern Africa, of modern humans before the Upper Paleolithic expansion all across Eurasia, as proposed by

the fossil evidence⁵⁵ and other genetic data.⁵⁶ The lack of basal limbs in Central Asian samples and the presence of lineages belonging to external branches within the mtDNA phylogeny suggest that the mtDNA diversity found in Africa did not have its 'maturation phase' in Central Asia, and the diversity found in the region is mainly the result of admixture of already differentiated populations. The lack of mtDNA basal root types in Central Asia contrasts with the results of Y-chromosome analyses. Whereas the majority of extant Y lineages in Europe and Siberia appear to have expanded from the Middle East via Central Asia,⁸ the lack of deeply rooting mtDNA clades in Central Asia does not support the hypothesis that Central Asia is the maternal source population for the Upper Paleolithic colonization of Europe. This discrepancy might be the result of different sexual migration patterns in Central Asia, as noted above. Additional data from autosomal markers, such as SNP or SNPSTR haplotypes,⁵⁷ need to be gathered in order to clarify the genetic role of Central Asia in the settlement of modern humans in Europe and Siberia.

Acknowledgements

Some individual European and West Asian sequences (published as haplogroups) were kindly provided by Martin Richards, Huddersfield University, UK. We thank Miguel A Padilla and Mònica Vallés, Universitat Pompeu Fabra, for technical assistance. The present study was supported by the Dirección General de Investigación Científica y Técnica, Spain (BOS2001-0794), and Direcció General de Recerca, Generalitat de Catalunya (2001SGR00285). SP received a fellowship from the Direcció General de Recerca, Generalitat de Catalunya (2000FI00696).

References

- 1 Okladnikov AP: Nakhodka Neandertal'tsa v Uzbekistane. *Vestnik Drevnei Istorii* 1939; 1: 256–257.
- 2 Derevyanko AP, Zun-E L: Upper Palaeolithic cultures. in Danni AH, Masson VM (eds) *History of the Civilizations of Central Asia*. Paris: UNESCO; 1992, Vol I.
- 3 Comas D, Calafell F, Mateu E *et al*: Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet* 1998; 63: 1824–1838.
- 4 Wang L, Oota H, Saitou N, Jin F, Matsushita T, Ueda S: Genetic structure of a 2500-year-old human population in China and its spatiotemporal changes. *Mol Biol Evol* 2000; 17: 1396–1400.
- 5 Bowles GT: *The People of Asia*. Birkenhead, Great Britain: Willmer Brothers Limited; 1977.
- 6 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton: Princeton University Press; 1994.
- 7 Pérez-Lezaun A, Calafell F, Comas D *et al*: Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am J Hum Genet* 1999; 65: 208–219.
- 8 Wells RS, Yuldasheva N, Ruzibakiev R *et al*: The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 2001; 98: 10244–10249.
- 9 Karafet T, Xu L, Du R *et al*: Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 2001; 69: 615–628.
- 10 Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C: A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet* 2002; 71: 466–482.

- 11 Ingman M, Kaessmann H, Paabo S, Gyllensten U: Mitochondrial genome variation and the origin of modern humans. *Nature* 2000; **408**: 708–713.
- 12 Finnila S, Lehtonen MS, Majamaa K: Phylogenetic network for European mtDNA. *Am J Hum Genet* 2001; **68**: 1475–1484.
- 13 Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM: Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2001; **2**: 13–20.
- 14 Herrnstadt C, Elson JL, Fahy E *et al*: Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 2002; **70**: 1152–1171.
- 15 Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P: Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 2003; **73**: 671–676.
- 16 Macaulay V, Richards M, Hickey E *et al*: The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999; **64**: 232–249.
- 17 Kivisild T, Tolk HV, Parik J *et al*: The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 2002; **19**: 1737–1751.
- 18 Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC: Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci USA* 1989; **86**: 9350–9354.
- 19 Anderson S, Bankier AT, Barrell BG *et al*: Sequence and organization of the human mitochondrial genome. *Nature* 1981; **290**: 457–465.
- 20 Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999; **23**: 147.
- 21 Shields GF, Schmiechen AM, Frazier BL *et al*: MtDNA sequences suggest a recent evolutionary divergence for Beringian and Northern North American populations. *Am J Hum Genet* 1993; **53**: 549–562.
- 22 Kolman CJ, Sambuughin N, Bermingham E: Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 1996; **142**: 1321–1334.
- 23 Kong QP, Yao YG, Liu M *et al*: Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet* 2003; **113**: 391–405.
- 24 Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L: From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 1996; **60**: 35–49.
- 25 Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bertranpetit J: Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 1996; **13**: 1067–1077.
- 26 Richards M, C rte-Real H, Forster P *et al*: Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 1996; **59**: 185–203.
- 27 Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP: Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 2002; **70**: 635–651.
- 28 Horai S, Murayama K, Hayasaka K *et al*: MtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am J Hum Genet* 1996; **59**: 579–590.
- 29 Lee SD, Shin CH, Kim KB, Lee YS, Lee JB: Sequence variation of mitochondrial DNA control region in Koreans. *For Sci Int* 1997; **87**: 99–116.
- 30 Koyama H, Iwasa M, Maeno Y *et al*: Mitochondrial sequence haplotype in the Japanese population. *For Sci Int* 2002; **125**: 93–96.
- 31 Richards M, Macaulay V, Hickey E *et al*: Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000; **67**: 1251–1276.
- 32 Comas D, Calafell F, Bendukidze N, Fa nan s L, Bertranpetit J: Georgian and Kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am J Phys Anthropol* 2000; **112**: 5–16.
- 33 Nasidze I, Stoneking M: Mitochondrial DNA variation and language replacements in the Caucasus. *Proc R Soc Lond B Biol Sci* 2001; **268**: 1197–1206.
- 34 Fucharoen G, Fucharoen S, Horai S: Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet* 2001; **46**: 115–125.
- 35 Kivisild T, Bamshad MJ, Kaldma K *et al*: Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 1999; **9**: 1331–1334.
- 36 Orekhov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov P, Yankovsky N: Mitochondrial DNA sequence diversity in Russians. *FEBS Lett* 1999; **445**: 197–201.
- 37 Malyarchuk BA, Derenko MV: Mitochondrial DNA variability in Russians and Ukrainians: implication to the origin of the Eastern Slavs. *Ann Hum Genet* 2001; **65**: 63–78.
- 38 Derenko MA, Gzybowski T, Malyarchuk A *et al*: Diversity of mitochondrial lineages in South Siberia. *Ann Hum Genet* 2003; **67**: 391–411.
- 39 Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI: Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 2002; **70**: 1009–1014.
- 40 Bandelt HJ, Forster P, Sykes BC, Richards MB: Mitochondrial portraits of human populations using median networks. *Genetics* 1995; **141**: 743–753.
- 41 Saillard J, Magalhaes PJ, Schwartz M, Rosenberg T, Norby S: Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. *Hum Biol* 2000; **72**: 1065–1068.
- 42 Dupanloup I, Bertorelle G: Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol* 2001; **18**: 672–675.
- 43 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.
- 44 Schneider S, Kueffer JM, Roessli D, Excoffier L: *Arlequin (ver.1.0): A Software Environment for the Analysis of Population Genetics Data*. University of Geneva, Switzerland: Genetics and Biometry Lab; 1996.
- 45 Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC: Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 1995; **57**: 133–149.
- 46 Watson E, Forster P, Richards M, Bandelt HJ: Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 1997; **61**: 691–704.
- 47 Rando JC, Pinto F, Gonzalez AM *et al*: Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* 1998; **62**: 531–550.
- 48 Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS: Genetic evidence of an early exit of Homo sapiens from Africa through eastern Africa. *Nat Genet* 1999; **23**: 437–441.
- 49 Torroni A, Bandelt HJ, D’Urbano L *et al*: MtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 1998; **62**: 1137–1152.
- 50 Plaza S, Calafell F, Lefranc F, Helal A, Bertranpetit J, Comas D: Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean. *Ann Hum Genet* 2003; **67**: 312–328.
- 51 Bamshad M, Kivisild T, Watkins WS *et al*: Genetic evidence on the origins of Indian caste populations. *Genome Res* 2001; **11**: 994–1004.
- 52 Oota H, Saitou N, Matsushita T, Ueda S: Molecular genetic analysis of remains of a 2000-year-old human population in China – and its relevance for the origin of the modern Japanese population. *Am J Hum Genet* 1999; **64**: 250–258.

- 53 Yao YG, Kong QP, Man XY, Bandelt HJ, Zhang YP: Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol Biol Evol* 2003; **20**: 214–219.
- 54 Seielstad MT, Minch E, Cavalli-Sforza LL: Genetic evidence for a higher female migration rate in humans. *Nat Genet* 1998; **20**: 278–280.
- 55 Stringer CB: The origin of early modern humans: a comparison of the European and non-European evidence. in Mellars P, Stringer CB (eds) *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*. Princeton: Princeton University Press; 1989, pp 232–244.
- 56 Forster P, Torroni A, Renfrew C, Röhl A: Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 2001; **18**: 1864–1881.
- 57 Mountain JL, Knight A, Jobin M *et al*: SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res* 2002; **12**: 1766–1772.