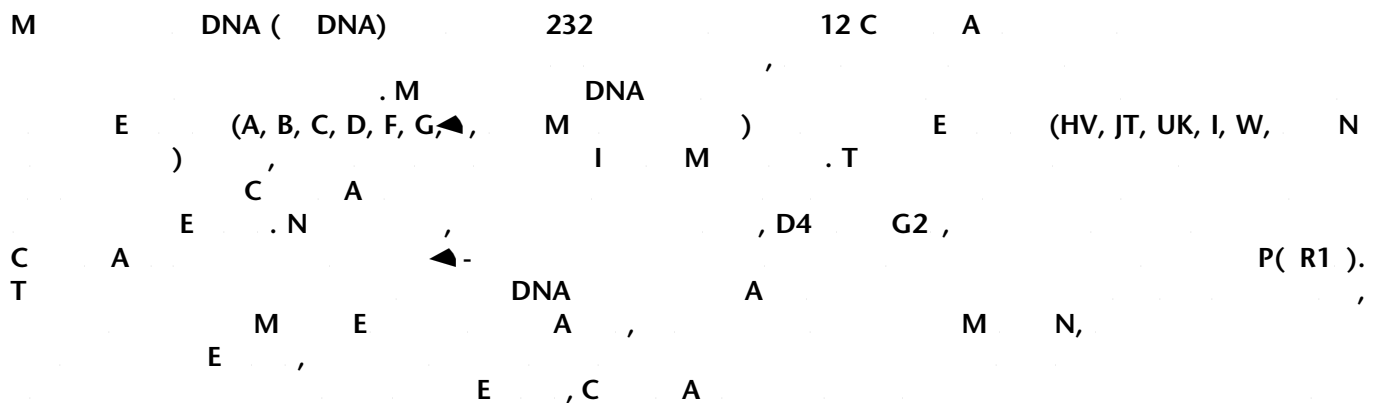


# Autosomal DNA in Central Asia

David Comas<sup>\*,1,4</sup>, Stéphanie Plaza<sup>1,4</sup>, R. Spencer Wells<sup>2</sup>, Nadira Yuldaseva<sup>2,3</sup>, Oscar Lao<sup>1</sup>, Francesc Calafell<sup>1</sup> and Jaume Bertranpetit<sup>1</sup>

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



*European Journal of Human Genetics* a a c a b ca , 11 Feb a 2004; :10.1038/ .5201160

K : m c a DNA; c a m ; C a A a

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,<sup>1</sup> Uzbekistan. Nonetheless, the later expansion of Upper Paleolithic remains is far less clear.<sup>2</sup> 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 descrip-C/, of west



16024–16391 and 63–322, respectively,<sup>19,20</sup> were obtained.

The 9-bp tandem repeat (CCCCCTCTA) of the COII/tRNA

Program Admix 2.0<sup>42</sup> 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 Europeans<sup>31</sup> (Bulgarians, Romanians, and Russians), 316 Middle Easterns<sup>31</sup> (Bedouins, Syrians, and Turks), 190

Northern Indians, and Pakistanis<sup>35</sup> (regions of Uttar Pradesh, Rajasthan, Punjab, Kashmir, Haryana, and Pakistan), and 263 East Asians<sup>27</sup> (Han Chinese).

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

T 1 Ha . . . a a a .  
m

---

---

---

---

## **R**

### **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<sup>Lys</sup> 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

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

An admixture approach<sup>42</sup>

accounting for a  $\sim 8\%$  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

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 China<sup>53</sup> 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<sup>17</sup>) 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<sup>17</sup> 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,<sup>9</sup> 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.<sup>9</sup>

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,<sup>10</sup> 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,<sup>7</sup> and as a global trend in which higher female than male migration has been observed.<sup>54</sup>

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'<sup>17</sup>), and a clear phylogeography in more external branches (haplogroups or sub-haplogroups; 'twigs'<sup>17</sup>) 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 evidence<sup>55</sup> and other genetic data.<sup>56</sup> 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,<sup>8</sup> 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,<sup>57</sup> need to be gathered in order to clarify the genetic role of Central Asia in the settlement of modern humans in Europe and Siberia.

#### A

Some individual European and West Asian sequences (published as haplogroups) were kindly provided by Martin Richards, Huddersfield





- 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