

## Origin of Caucasoid-Specific Mitochondrial DNA Lineages in the Ethnic Groups of the Altai–Sayan Region

M. V. Derenko<sup>1</sup>, B. A. Malyarchuk<sup>1</sup>, and I. A. Zakharov<sup>2</sup>

<sup>1</sup> *Institute of Biological Problems of the North, Russian Academy of Sciences, Magadan, 685000 Russia;*  
fax: (41322)344-63; e-mail: mderenko@mail.ru

<sup>2</sup> *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia*

Received June 26, 2001; in final form, January 9, 2002

**Abstract**—The data on sequence variation in the first hypervariable segment (HVSI) of human mitochondrial DNA (mtDNA) representing Caucasoid mtDNA lineages in the gene pools of Altaians and Khakassians are presented. Identification of the subgroups of Caucasoid mtDNA lineages found in the gene pools of the ethnic groups of the Altai–Sayan region and the adjacent territories, Altaians, Khakassians, Tuvinians, Buryats, and Yakuts was carried out. All Caucasoid mtDNA lineages belonged to groups H, HV1, J\*, J1, J1b1, T1, T4, U1a, U2, U3, U4, U5a1, I, X and N1a. Taking into consideration possible contribution of southern Caucasoid and eastern European components to the formation of the anthropological type of Altai–Sayan ethnic groups, distribution of the revealed Caucasoid mtDNA lineages among the ethnic groups of the Central Asia, Western Asia, Caucasus, and Eastern Europe was examined. The applied approach permitted identification of 60% of mtDNA types the majority of which had southern Caucasoid origin. Less than 10% of mtDNA types were of eastern European origin. The gene pools of Altaians and Khakassians displayed the presence of autochthonous components represented by mtDNA types from subgroups U2 and U4.

### INTRODUCTION

Ethnic history of populations from the Altai–Sayan region occurred on the territories which from the ancient times were the areas of intense admixture of Mongoloid and Caucasoid groups. According to paleoanthropological data, the Caucasoid (in respect of its morphological features) population predominated in the steppes of the Altai–Sayan region during the Neolithic, Bronze, and partly early Iron Ages [1–3]. At that time the Mongoloid component was observed only in few cases. However, beginning from the early Iron Age, the presence of this component has been increasing, and becoming prevalent in modern times. Thus, dynamics of the anthropological composition of the Altai–Sayan populations can be characterized by definitely directed replacement of the Caucasoid component by the Mongoloid one. It should be noted in this respect that this process was nonuniform. Because of this, certain forms and the chronology of the admixture of Caucasoid and Mongoloid populations in the region studied were different. Anthropological data demonstrate the territorial differences in the chronology of ethnogenetic processes in at least three largest populations of the Altai–Sayan region: Khakassians, Altaians, and Tuvinians. Most intense formation of Khakassian population dates back to the end of the first millennium C.E., while the admixture of Mongoloids and Caucasoids in the Altai was finished at the boundary between the first and second millenniums C.E. In Tuva the prevalence of Caucasoids can be followed up to the pre-Mongolian time [2]. Moreover, modern Tuvinians dis-

play anthropological traits specific for the ancient population characterized by southern Caucasoid traits. On the other hand, Khakassians and Shorians demonstrate the influence of Eastern European traits [2, 3].

Multicomponent composition of the Altai–Sayan gene pools is confirmed by the data of molecular genetics. Analysis of variation in highly polymorphic genetic systems, which are inherited in one of the parental lineages, mitochondrial DNA (mtDNA) and the Y chromosome, showed that Caucasoid components with different frequencies could be found in the gene pools of all ethnic groups of the Altai–Sayan region and adjacent territories examined [4, 5]. Minimum frequencies of the Caucasoid lineages are typical of Tuvinians and Buryats, while in other populations of the region, an increase of the proportion of the Caucasoid component with its maximum in Shorians and Southern Altaians is observed [4].

The data on the structure of mitochondrial gene pools of the Altai–Sayan ethnic groups obtained point to the heterogeneous nature of Caucasoid component represented by mtDNA types belonging to groups H, U, K, T, J, and I, specific to Caucasoid populations [4, 6]. Here we present a detailed description (at the level of nucleotide sequences of the first hypervariable segment (HVSI)) of Caucasoid mtDNA lineages revealed in the gene pools of the ethnic groups of the Altai–Sayan region and adjacent territories. Using the phylogeographic approach, it was possible to identify 60% of Caucasoid mtDNA lineages and make conclu-

sion on the southern Caucasoid ancestry of most of them.

## MATERIALS AND METHODS

Nucleotide sequences of the mtDNA HVSI between nucleotide positions 16000 and 16400 (according to the numeration of the Cambridge reference sequence [7]) were determined in the sample group characterized on the basis of the restriction analysis data [4, 6] by the presence of Caucasoid mtDNA types, as well as by the types not included into Mongoloid mtDNA groups, i.e., A, B, F, M\*, C, D, E, and G. The sample group was formed of Altaians (41 individuals, the inhabitants of different regions of the Altai Republic), and Khakassians (8 individuals, the inhabitants of Askizsk, Shirinsk, Beisk, and Ordzhonikidze raions of the Khakassia Republic).

DNA was extracted from the samples of biological tissues (blood and hair bulbs) using standard technique. Polymerase chain reaction and DNA sequencing were performed as described earlier [6].

In this study the data on the mtDNA HVSI sequences in the populations of Tuvinians, Buryats and Yakuts reported by us earlier [6, 8], as well as the literature data (for the list of publications, see [9–11]) on the sequence variation in this mtDNA region in other populations of Eurasia (the database of more than 5000 HVSI sequences) were used. In order to identify Caucasoid mtDNA types found in the populations of the Altai–Sayan region a search for phylogenetically related mtDNA types (identical or derivative sequences differing by 1 or 3 nucleotide substitutions) was carried out. The search was performed using the data for the populations inhabiting those regions which, according to the anthropological data, are the places from where the peopling of Southern Siberia began. These regions are as follows: (1) Central Asia (Kyrgyzes, Kazakhs, and Uigurs ( $N = 205$ )); and (2) Eastern Europe (Russians and the populations of Volga–Ural region, mainly represented by Finno-Ugrians ( $N = 421$ )). In order to obtain more detailed information on the diversity of mtDNA lineages in the gene pools of the populations belonging to the southern Caucasoid branch, the data on the population of Western Asia, represented by Turks and Iranians, and the population of the Caucasus, mainly represented by Transcaucasian populations (Azerbaijanians, Armenians, and Kurds with the pooled sample size of 544 individuals) were utilized.

Identification of Caucasoid mtDNA types was carried out based on the data on restriction polymorphism variants determining mtDNA groups and on the HVSI motifs in accordance with the classification of Eurasian mtDNA types suggested in [10].

Statistical significance of the differences between the regions in respect of the mtDNA lineage frequencies was estimated using  $\chi^2$  test.

## RESULTS AND DISCUSSION

### *Structure of Causoid-Specific mtDNA Lineages in the Gene Pools of the Ethnic Groups of the Altai–Sayan Region and Adjacent Territories*

Sequences of mtDNA HVSI were examined in the population sample of the Altai–Sayan and neighboring ethnic groups (Altaians, Khakassians, Tuvinians, Buryats, and Yakuts), whose mtDNA types, according to the data of restriction analysis, belonged to either Caucasoid groups (H, HV, J, T, U, I, and X), or to the category of “other” mtDNA types unidentified in our previous study [4]. The results of identification of mtDNA samples carried out in accordance with the classification of Eurasian mtDNAs are presented in Table 1. Caucasoid-specific mtDNA component of the populations examined was characterized by substantial diversity at the level of mtDNA subgroups. Maximum diversity was observed in the Altaian gene pool of, where mtDNA types belonging to five subgroups of group H, five subgroups of group U, and also single mtDNA types from groups J, T, I, and X were identified. In Altaians, mtDNA types belonging to different groups of macrogroup N were also found. In addition to N1a mtDNA types (Table 1), distributed predominantly in Caucasoid populations [10], Altaians and Khakassians demonstrated the presence of mtDNA types of subgroup N\* (its nucleotide motif is 16223, 16257CA, 16261) typical of Mongoloid populations [10]. Among previously unidentified by restriction analysis mtDNA types, the HVSI sequences characterized by nucleotide motif 16092TA, 16291, 16304 and belonging to group R\* were detected in Altaians and Yakuts. The types of mtDNA from this rare group are mainly observed in Mongoloid populations (Mongols, Chinese, and Evens) [8, 12, 13].

In Khakassians, which share mtDNA types with Altaians (H and U4), single types belonging to subgroups J1b1 and T4 were also identified. Minimum number of Caucasoid-specific lineages was found in Tuvinians and Buryats (Table 1). It should be noted that these mtDNA types belong to different subgroups. Thus, Caucasoid-specific lineages in the gene pools of the populations of the Altai–Sayan region and adjacent territories are characterized by substantial heterogeneity. It was shown that the types of mtDNA belonging to groups U, H, and J were most common among the Altai–Sayan populations, accounting for 40.4, 25 and 13.5% of all sequences in the pooled Caucasoid component respectively.

### *Origin of Caucasoid-Specific Mitochondrial DNA Lineages in the Ethnic Groups of Altai–Sayan Region and Adjacent Territories*

In order to reveal the origin of mtDNA types found in the populations of Altai–Sayan region and adjacent territories, an analysis of their distribution among other populations of Eurasia was carried out. In case of

**Table 1.** Distribution of mtDNA HVSI types among the populations of Southern Siberia

Groups of mtDNA and nucleotide sequence of mtDNA types	Altaians <i>N</i> = 110	Khakas- sians <i>N</i> = 42	Tuvinians <i>N</i> = 36	Buryats <i>N</i> = 40	Yakuts <i>N</i> = 22
<b>H:</b>					
H CRS			1		
H 16271				1	
H 16223				1	
H 16288, 16362	1	2			
H 16362					1
H 16169, 16184	2				
H 16311	1				
H 16354	2				
H 16092, 16245, 16362	1				
<b>HV:</b>					
HV1 16067, 16248, 16260, 16355					1
<b>J:</b>					
J* 16069, 16126					2
J1 16069, 16126, 16145, 16261, 16290	4				
J1b1 16069, 16126, 16145, 16172, 16222, 16261		1			
<b>T:</b>					
T4 16126, 16168, 16294, 16296, 16324		1			
T1 16126, 16163, 16186, 16189, 16294	1				
<b>U:</b>					
U1a 16192, 16249, 16311	3				
U2 16129GC, 16189, 16362			1		
U2 16189, 16214, 16362	1				
U2 16051, 16129GC, 16189, 16214, 16258, 16362	4				
U3 16343	2				
U4 16356	4	1			
U4 16311, 16356	2	2			
U5a1 16192, 16241, 16256, 16270, 16287, 16304, 16325, 16399	1				
<b>I:</b>					
I 16129, 16223, 16391	2				
<b>X:</b>					
X 16189, 16223, 16278	3				
<b>N:</b>					
N1a 16147CA, 16172, 16189, 16223, 16248, 16320, 16355	2				
N1a 16147CG, 16172, 16189, 16223, 16248, 16320, 16355	1				
N* 16189, 16223, 16257CA, 16261	2				
N* 16223, 16248, 16257CA, 16261, 16311	3				
N* 16172, 16223, 16257CA, 16261		1			
<b>R*:</b>					
R* 16092TA, 16291, 16304	1				1

Note: In the HVSI sequences the positions of transitions differing this mtDNA type from the Cambridge Reference Sequence (CRS) [7] are shown. *N*, sample size.

**Table 2.** Regional specificity and the origin of the Caucasoid-specific mtDNA lineages found in the populations of Altai--Sayan region

Groups of mtDNA and nucleotide sequence of mtDNA types	Altai-Sayan region ( <i>N</i> = 250)	Central Asia ( <i>N</i> = 205)	Caucasus and Western Asia ( <i>N</i> = 544)	Eastern Europe ( <i>N</i> = 421)	Origin of mtDNA types
H: CRS	1 (0.4)	11 (5.37)	61 (11.21)	57 (13.54)	Uncertain
H: 16271	1 (0.4)	0	1 (0.18)	1 (0.24)	Uncertain
H: 16223	1 (0.4)	0	1 (0.18)	2 (0.48)	Uncertain
H: 16288, 16362	3 (1.2)	1 (0.49)	2 (0.37)	0	southern-Caucasoid (*)
H: 16362	1 (0.4)	0	4 (0.74)	10 (2.38)	eastern-European (**)
H: 16169, 16184	2 (0.8)	0	0	1 (0.24)	eastern-European
H: 16311	1 (0.4)	2 (0.98)	4 (0.74)	5 (1.19)	Uncertain
H: 16354	2 (0.8)	4 (1.95)	7 (1.29)	7 (1.66)	southern-Caucasoid
H: 16092, 16245, 16362	1 (0.4)	1 (0.49)	3 (0.55)	0	southern-Caucasoid (*)
HV1: 16067, 16248, 16260, 16355	1 (0.4)	2 (0.98)	4 (0.74)	0	southern-Caucasoid (**)
J*: 16069, 16126	2 (0.8)	0	13 (2.39)	23 (5.46)	Uncertain (***)
J1: 16069, 16126, 16145, 16261, 16290	4 (1.6)	0	7 (1.29)	0	southern-Caucasoid (***)
J1b1: 16069, 16126, 16145, 16172, 16222, 16261	1 (0.4)	1 (0.49)	2 (0.37)	0	southern-Caucasoid (*)
T4: 16126, 16168, 16294, 16296, 16324	1 (0.4)	0	5 (0.92)	1 (0.24)	southern-Caucasoid (*)
T1: 16126, 16163, 16186, 16189, 16294	1 (0.4)	0	9 (1.65)	8 (1.9)	Uncertain
U1a: 16192, 16249, 16311	3 (1.2)	3 (1.46)	8 (1.47)	3 (0.71)	southern-Caucasoid
U2: 16129GC, 16189, 16362	1 (0.4)	1 (0.49)	5 (0.92)	7 (1.66)	Uncertain
U2: 16189, 16214, 16362	1 (0.4)	0	1 (0.18)	0	Southern Siberian
U2: 16051, 16129GC, 16189, 16214, 16258, 16362	4 (1.6)	0	1 (0.18)	0	Southern Siberian
U3: 16343	2 (0.8)	0	21 (3.86)	1 (0.24)	southern-Caucasoid (***)
U4: 16356	5 (2.0)	4 (1.95)	17 (3.13)	22 (5.23)	Uncertain (**)
U4: 16311, 16356	4 (1.6)	0	0	0	Southern Siberian
U5a1: 16192, 16241, 16256, 16270, 16287, 16304, 16325, 16399	1 (0.4)	0	2 (0.37)	3 (0.71)	Uncertain
I: 16129, 16223, 16391	2 (0.8)	4 (1.95)	12 (2.21)	3 (0.71)	southern-Caucasoid (**)
X: 16189, 16223, 16278	3 (1.2)	0	12 (2.21)	1 (0.24)	southern-Caucasoid (***)
N1a: 16147CA, 16172, 16189, 16223, 16248, 16320, 16355	2 (0.8)	0	1 (0.18)	2 (0.48)	Uncertain
N1a: 16147CG, 16172, 16189, 16223, 16248, 16320, 16355	1 (0.4)	0	1 (0.18)	2 (0.48)	Uncertain

Note: Population frequencies (%) of mtDNA lineages are in brackets. *N*, sample size. Statistical significance of the differences in the frequencies of mtDNA lineages was determined between the populations of Eastern Europe and the populations of Central Asia, Western Asia, and the Caucasus. \*  $P < 0.1$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

absence of identical mtDNA types a search for closer by the origin mtDNA types, differing from the tested ones by the minimum number of mutations, was performed. Table 2 presents the data on the distribution of phylogenetically related mtDNA types among the populations of Southern Siberia and two other regions, Central Asia and Eastern Europe, from where, according to anthropological data, the populations represent-

ing southern (Central Asia) and eastern-European branches of Caucasoids migrated to Southern Siberia [1–3]. Given that the mean proportion of the Caucasoid-specific component in the gene pool of Central Asian populations constitutes only 30.5% (other lineages are of the Mongoloid descent) [14], the populations in which the complex of anthropological traits typical of Southern Caucasoids was best represented,

i.e., the populations of Western Asia (Iranians and Turks) and the Caucasus (mainly Transcaucasian populations), were also included in the analysis. As seen from Table 2, all HVSI sequences examined could be classified into four categories, namely, those having southern Caucasoid, Eastern European, Southern Siberian, and uncertain descent. The last category includes either mtDNA types widely distributed in different Caucasoid populations (like, for example, most ancient mtDNA subgroups and groups founder types) or rare mtDNA types, whose geographic frequency distribution patterns give no evidence on their origin. Of 27 examined mtDNA types constituting Caucasoid component in the gene pools of Altai–Sayan populations, 11 had an uncertain origin. The remaining mtDNA types were characterized by regional specificity at the level of the subgroups distribution. The frequency distribution of mitochondrial lineages in the populations of the Eurasian regions compared indicated that virtually all mtDNA types revealed in the Altai–Sayan populations were of the southern Caucasoid origin. These mtDNA types accounted for 40.7% of all sequences in the gene pools of the populations studied. Only 7.4% of mtDNA types could be of the Eastern European origin (both of these types belonged to group H). Three mtDNA types from group U did not have analogs in the populations of Eurasia. Because of this, it can be suggested that they had local (Southern Siberian) origin. The types of mtDNA belonging to subgroup U2 and characterized by marker variant 16214T can serve as a good illustration of this statement. A search for analogs of this mtDNA lineage in Eurasian populations showed that the variants of these U2-specific mtDNA types were distributed only among Altaians with the frequency of 4.7% ([5] and the present study). A single sequence of this type was found in the Caucasus (Adygeis, 2% [16]).

Our data show that the ancestry of most of Caucasoid-specific mtDNA types (more than 40%) found in the population of Southern Siberia is associated with the populations of Southern Caucasoids. In view of the existence of ancient genetic relationships between the populations inhabiting the south of Siberia and Central Asia, it can be suggested that many of unidentified mtDNA types (with uncertain origin, Table 2) also appeared due to the migrations of Southern Caucasoids to Siberia. In this case, the genetic relatedness between the population of Southern Siberia and Central Asia must become closer. In any case, the need for more detailed analysis of the relatedness between the populations of Siberia and Central Asia requires a further increase of the number of samples from these regions. The new data are also needed for the estimation of the evolutionary age of the Caucasoid-specific lineages belonging to the gene pools of the ethnic groups of Southern Siberia and neighboring territories. Based on the data of the present study it can be concluded that Caucasoid component in the populations of Southern Siberia is characterized by substantial diversity, which

is most expressed in the Altai–Sayan populations. The existence of autochthonous mtDNA lineages, like that of U2 type with the marker variant 16214T, suggests that evolution of some of Caucasoid-specific mtDNA lineages in this region occurred during many thousand years. For example, the evolutionary age of the U2-214T mtDNA lineage is about 30 000 years. This means that further accumulation of the data on the mtDNA diversity in the populations of Altai–Sayan region will shed the light on the involvement of ancient Caucasoids in the formation of the genetic pattern of Siberian populations. These data are important not only for understanding the genetic history of the Siberian populations, but also for elucidating the genetic relationships between the Eurasian ethnic groups.

#### ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grant no. 99-06-80430) and the Russian State Program “Frontiers in Genetics” (grant no. 99-04-30).

#### REFERENCES

1. Alekseev, V.P. and Gokhman, I.I., *Anthropology of the Asian Part of the USSR*, Moscow: Nauka, 1984.
2. Alekseev, V.P., *Istoricheskaya antropologiya i antropogenez* (Historical Anthropology and Anthropogenesis), Moscow: Nauka, 1989.
3. *Antropo-ekologicheskie issledovaniya v Tuve* (Anthropo-Ecological Studies in Tuva), Alekseeva, T.I. and Uryson, M.I., Eds., Moscow: Nauka, 1984.
4. Derenko, M.V., Denisova, G.A., Malyarchuk, B.A., *et al.*, The Gene Pool Structure of Ethnic Groups of the Altai–Sayan Highlands as Inferred from the Data on Mitochondrial DNA Polymorphism, *Genetika* (Moscow), 2001, vol. 37, no. 10, pp. 1402–1410.
5. Derenko, M.V., Malyarchuk, B.A., Denisova, G.A., *et al.*, Polymorphism of Y-Chromosomal Diallelic Loci in the Indigenous Population of the Altai–Sayan Highlands, *Genetika* (Moscow), 2002, vol. 38, no. 3, pp. 393–399.
6. Derenko, M.V., Malyarchuk, B.A., Dambueva, I.K., *et al.*, Mitochondrial DNA Variation in Two South Siberian Aboriginal Populations: Implications for the Genetic History of North Asia, *Hum. Biol.*, 2000, vol. 72, pp. 945–973.
7. Anderson, S., Bankier, A.T., Barrel, B.G., *et al.*, Sequence and Organization of the Human Mitochondrial Genome, *Nature*, 1981, vol. 290, pp. 457–465.
8. Derenko, M.V. and Shields, J.F., Nucleotide Sequence Diversity of Mitochondrial DNA in Three Groups of the Indigenous Population of Northern Asia, *Mol. Biol.* (Moscow), 1997, vol. 31, no. 5, pp. 784–789.
9. Simoni, L., Calafell, F., Pettener, D., *et al.*, Geographic Patterns of mtDNA Diversity in Europe, *Am. J. Hum. Genet.*, 2000, vol. 66, pp. 262–278.
10. Richards, M., Macaulay, V., Hickey, E., *et al.*, Tracing European Founder Lineages in the Near Eastern

- mtDNA Pool, *Am. J. Hum. Genet.*, 2000, vol. 67, pp. 1251–1276.
11. Malyarchuk, B.A. and Derenko, M.V., Mitochondrial DNA Variability in Russians and Ukrainians: Implication to the Origin of the Eastern Slavs, *Ann. Hum. Genet.*, 2001, vol. 65, pp. 63–78.
  12. Kolman, C.J., Sambuughin, N., and Bermingham, E., Mitochondrial DNA Analysis of Mongolian Populations and Implications for the Origin of New World Founders, *Genetics*, 1996, vol. 142, pp. 1321–1337.
  13. 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, vol. 59, pp. 579–590.
  14. Kivisild, T., Kaldma, K., Metspalu, E., *et al.*, The Place of the Indian Mitochondrial DNA Variants in the Global Network of Maternal Lineages and the Peopling of the Old World, *Genomic Diversity: Applications in Human Population Genetics*, Papiha, S., Deka, R., and Chakraborty, R., Eds., New York: Kluwer Academic, 1999, pp. 135–152.
  15. Shields, G.F., Schmiechen, A.M., Frazier, B.L., *et al.*, mtDNA Sequences Suggest a Recent Evolutionary Divergence for Beringian and Northern North American Populations, *Am. J. Hum. Genet.*, 1993, vol. 53, pp. 549–562.
  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, vol. 64, pp. 232–249.