
**GENOMICS. TRANSCRIPTOMICS.
PROTEOMICS**

UDC 575.174.599.9

Diversity of Mitochondrial DNA Haplogroups in Ethnic Populations of the Volga–Ural Region

M. A. Bermisheva¹, K. Tambets², R. Villems², and E. K. Khusnutdinova¹

¹ *Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa, 450054 Russia;
E-mail: ekkh@anrb.ru*

² *Estonian Biological Center, Tartu University, Tartu, Estonia*

Received February 28, 2002

Abstract—The mtDNA polymorphism was analyzed in eight ethnic groups ($N = 979$) of the Volga–Ural region. Most mtDNA variants belonged to haplogroups H, U, T, J, W, I, R, and N1 characteristic of West Eurasian populations. The most frequent were haplogroups H (12–42%) and U (18–44%). East Eurasian mtDNA types (A, B, Y, F, M, N9) were also observed. Genetic diversity was higher in Turkic than in Finno-Ugric populations. The frequency of mtDNA types characteristic of Siberian and Central Asian populations substantially increased in the ethnic groups living closer to the Urals, a boundary between Europe and Asia. Geographic distances, rather than linguistic barriers, were assumed to play the major role in distribution of mtDNA types in the Volga–Ural region. Thus, as concerns the maternal lineage, the Finno-Ugric populations of the region proved to be more similar to their Turkic neighbors rather than to linguistically related Balto-Finnish ethnic groups.

Key words: mitochondrial DNA, haplotyping, ethnogenomics, ethnic groups of the Volga–Ural region

INTRODUCTION

Mitochondrial (mt) DNA and the Y chromosome occupy a specific place among highly polymorphic genetic systems. Owing to the maternal inheritance, absence of recombination, and high polymorphism, mtDNA can be employed as a major tool in evolutionary, populational, and medical genetic studies [1, 2]. The variation of human mtDNA has been extensively studied in the 1980s and 1990s. Race- and population-specific mtDNA types and mutations have been described and used to define individual haplogroups [2, 3]. A modern classification of European mtDNA variants is based on the combined data [4–6] on polymorphisms of the control (hypervariable segments I and II, HVS I and HVS II) [7] and coding [8] regions of mtDNA. The European mtDNA clusters are best organized owing to the African ancestor in cluster M, while haplogroups A and B originate from cluster N. Phylogenetically, various mtDNA variants specific for Eastern Eurasia may be classed into several macroclusters (Fig. 1). Macroclusters M and N originate from African macrocluster L3, the consensus nucleotide sequence of which is considered as a common ancestor (MRCA) of all non-African mtDNA lineages. Haplogroups of cluster M are frequent and diverse in Asia [10] and rare in Europe, which supports the hypothesis that Asia was first populated as a result of migration from Eastern Africa [11]. This cluster harbors about half mtDNA variants of the indigenous population of Eastern Eurasia [2], and consists of several haplogroups, including C, Z, D, G, E, M7, M8, M9, M10, etc. (Fig. 1). Supercluster N includes haplogroups A, Y, B, and F in Asia. Haplogroup A has an ancestral nucleotide (T) in

EXPERIMENTAL

Material for population genetic analysis was collected during expeditions from 1989 to 2001. Blood of informed adults was obtained after medical examination. In total, we tested 979 DNA specimens of the indigenous populations of the Volga–Ural region, including 221 Bashkirs (Trans-Ural, South-western, Northeastern, and Gaininsk Bashkirs) from the Beloretskii, Sterlibashevskii, and Ilishevskii Districts of Bashkortostan and the Perm Region; 228 Tatars from the Al'met'evskii and Elabuzhskii Districts of Tatarstan; 55 Chuvash from the Morgaushskii District of Chuvashia; 102 Mordvinians from the Staro-Shaiginskii District of Mordovia; 136 Mari from the Zvenigovskii District of Marii El; 101 Udmurts from the Malo-Purginskii District of Udmurtia and the Tatyshlinskii District of Bashkortostan; 62 Komi-Zyryans from the Sysol'skii District of the Komi Republic; and 74 Komi-Permyaks from the Komi-Permyak Autonomous District.

Genomic DNA was isolated from 10 ml of peripheral blood by phenol–chloroform extraction. Fragment 16,024–16,400 of mtDNA HVSI was sequenced in a Perkin-Elmer ABI 377 DNA sequencer with a DYEnamic ET terminator cycle sequencing premix kit (Amersham Pharmacia Biotech).

Haplogroups were identified by mtDNA RFLPs, including 73 (73GA) –*Alw*44I, 663 (663AG) +*Hae*III, 1715 (1719GA) –*Dde*I, 4577 (4580GA) –*Nla*III, 4643 (4646TC) +*Rsa*I, 4830 (4833AG) +*Hae*II, 5176 (5178CA) –*Alu*I, 5416 (5417 GA) +*Tas*I, 5656 (5656AG) +*Nhe*I, 7025 (7028CT) –*Alu*I, 7598 (7600GA) –*Hha*I, 8249 (8251CT) +*Ava*II, 8994 (8994GA) –*Hae*III, 9052 (9055GA) –*Hae*II, 10,032 (10,034TC) +*Alu*I, 10,394 (10,398AG) +*Dde*I, 10,397 (10,400CT) +*Alu*I, 12,308 (12,308AG) +*Hinf*I, 12,406 (12,406GA) –*Hinc*II, 12,704 (12,705CT) –*Mbo*II, 13,262 (13,263AG) +*Alu*I, 13,366 (13,368GA) +*Bam*HI, 13,704 (13,708GA) –*Bst*I, 14,068 (14,070AG) +*Taq*I, 14,766 (14,766TC) +*Mse*I, and 15,904 (15,904CT) +*Mse*I. Transition or transversion resulting in a gain or loss of a restriction site is indicated in parentheses. Mutations in mtDNA were identified according to the Cambridge reference sequence (CRS) [13]. To establish the phylogenetic relationships among haplotypes, median networks were constructed with the combined data on polymorphisms of the mtDNA coding region and HVSI for the total sample of the Volga–Ural region.

Genetic heterogeneity was calculated as

$$H = \left(\frac{\sum x_i^2}{N} \right) / N,$$

where x is the frequency of an mtDNA haplotype in a population and N is the sample size [14].

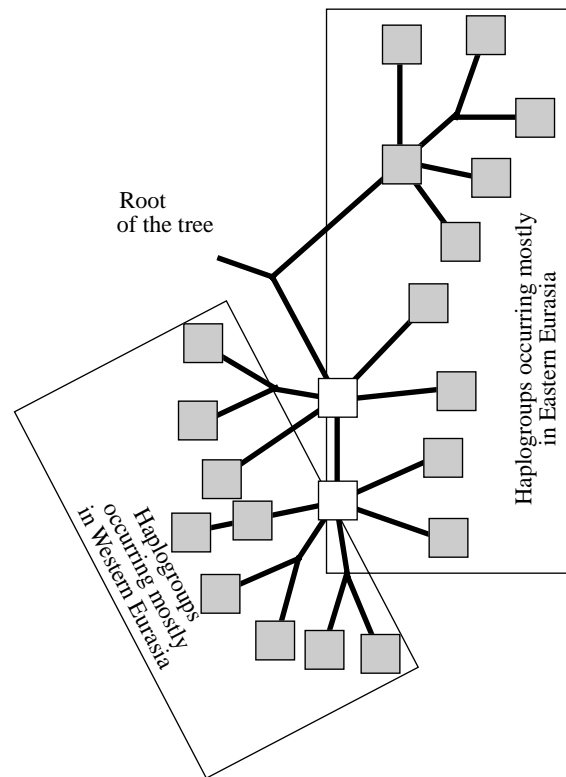


Fig. 1. Phylogenetic tree of mtDNA clusters found in Eurasian populations.

RESULTS AND DISCUSSION

The indigenous population of the Volga–Ural region is ethnically, historically, and culturally heterogeneous. Ethnic groups of the region belong to different language groups, including the Perm (Komi-Zyryans, Komi-Permyaks, Udmurts) and Volga-Finnish (Mordvinians, Mari) branches of the Ural language family and the Turkic branch of the Altaian language family (Bashkirs, Tatars, Chuvash). Anthropologically, these ethnic groups are Caucasian and have a varying Mongoloid component [12]. The region includes various geographic zones, from taiga, tundra (Komi-Zyryans) and Central Volga plains (Mordvinians, Mari, Chuvash, Tatars) to the Southern Urals and Trans-Ural steppe (Bashkirs) (Fig. 2).

We sequenced the 377-bp region (16,024–16,400) of mtDNA HVSI in 979 people of the eight ethnic groups of the Volga–Ural region. Haplogroups were identified by RFLP analysis of 26 polymorphic sites of the mtDNA coding region and analysis of the deletion from a region between the *COII* and *tRNA^{Lys}* genes. Genetic heterogeneity was relatively low (0.96) in Komi-Zyryans and Udmurts, and high (0.99) in Bashkirs. Mean heterogeneity of the Turkic ethnic groups (0.98) was higher than in the Finno-Ugric ethnic groups (0.96).

Haplogroup-determining HVSI motifs found in the Volga–Ural populations are shown in Table 1. On evidence of cluster analysis, most mtDNA types were assigned to haplogroups specific for West Eurasian populations (Table 2).

Specific mtDNA types have already been established for most race groups in the early 1990s [2, 4]. More than 80% mtDNA types of the West Eurasian populations belong to haplogroups H, I, J, T, U, V, W, and X, which are considered originating from the Upper Paleolithic European gene pool. According to published data, the population frequency of haplogroup H is maximal (40–50%) in Western and Northern Europe; intermediate (20–40%) in Southern, Southwestern, and Eastern Europe, Northern Africa, and Turkey; and low (less than 20%) in the Middle East, India, and Central Siberia [6]. This haplogroup

Mordvinians (about 42%) and Mari (about 40%) and

center occupied by the Cambridge sequence [3],

assumed from its geographic frequency distribution and genetic diversity, haplogroup V arose in Southwestern Europe about 16,000 years ago and spread

observed mtDNA types of this haplogroup in Bash-

Africa, while U7 is typical in the populations of Jordan, Kuwait, Iran, and Saudi Arabia. In addition, cluster U includes haplogroup K [4].

Haplogroups U5 and U4 were the most common in the Volga–Ural region. The former is prevalent (53.0%) in Saami and is also present in other North European populations [17–20]. In our sample, haplogroup U5 was not only observed in the Finno-Ugric populations, it also occurred at a high frequency in Bashkirs, Tatars, and Chuvash. As Fig. 4 shows, haplogroup U5 had several founder types. One of these, so-called Saami motif (16,144–16,189–16,270) has earlier been detected in Saami (37%), Finns (2%), and the Karelian population (6%) [20]. We found this motif in Mari, Mordvinians, and Bashkirs.

Of all cluster U haplogroups, U4 was the most common in Trans-Ural Bashkirs and Komi-Zyryans (16 and 24%, respectively). While haplogroup U5 is common in Western Europe and Mediterranean countries and occurs also in the Middle East and Central Asia, haplogroup U4 is characteristic of the northeastern populations of Western Europe. Haplogroup U4 was observed at a low frequency in Mordvinians (2%) and Udmurts (4%). We detected several founder variants with substitutions in positions 16,356, 16,356–16,261, 16,356–16,362–16,242A–16,288, and 16,356–16,134. Subcluster U4a (16,356–16,134) had a distinct stellate phylogenetic tree. The period of its divergence in the Volga–Ural region was estimated at $17,800 \pm 2900$ years ($\rho = 0.88$).

The populations of the Volga–Ural region proved to be heterogeneous with respect to haplogroup U2. This haplogroup is known to include two major variants. One (HVSI motif 16,051G, 16,129C, 16,189C) is specific for Europe, whereas the other is restricted to India [21]. European U2 lines were found in Bashkirs, Tatars, Mordvinians, and Udmurts.

Haplogroup U3 is rare in Europe [5] and absent from Central Asia [22]. A high genetic diversity of this haplogroup has been observed in the Caucasus (Osetia, Georgia, Armenia) and in Turkey [4, 23]. In our sample, Tatars and Chuvash had two variants, which have earlier been detected in Turks, Armenians, and Nogaians [23].

Variants of haplogroup U8 were rarely observed in Gaininsk Bashkirs, Komi-Zyryans, and Chuvash.

The distribution of haplogroup K in the populations of the Volga–Ural region is also interesting (Table 2). This haplogroup, which includes several mtDNA variants, was relatively frequent in the Turkic ethnic groups and was rarely, if at all, observed in the Finno-Ugric populations.

Haplogroups T and J originate from the Middle East [5, 7]. The former accounts for about 8% mtDNA lineages in Europe and includes two subclusters [5, 24]. In our sample, the frequency of haplogroup T was

Table 1. Haplogroup-determining mutations in mtDNA HVSI

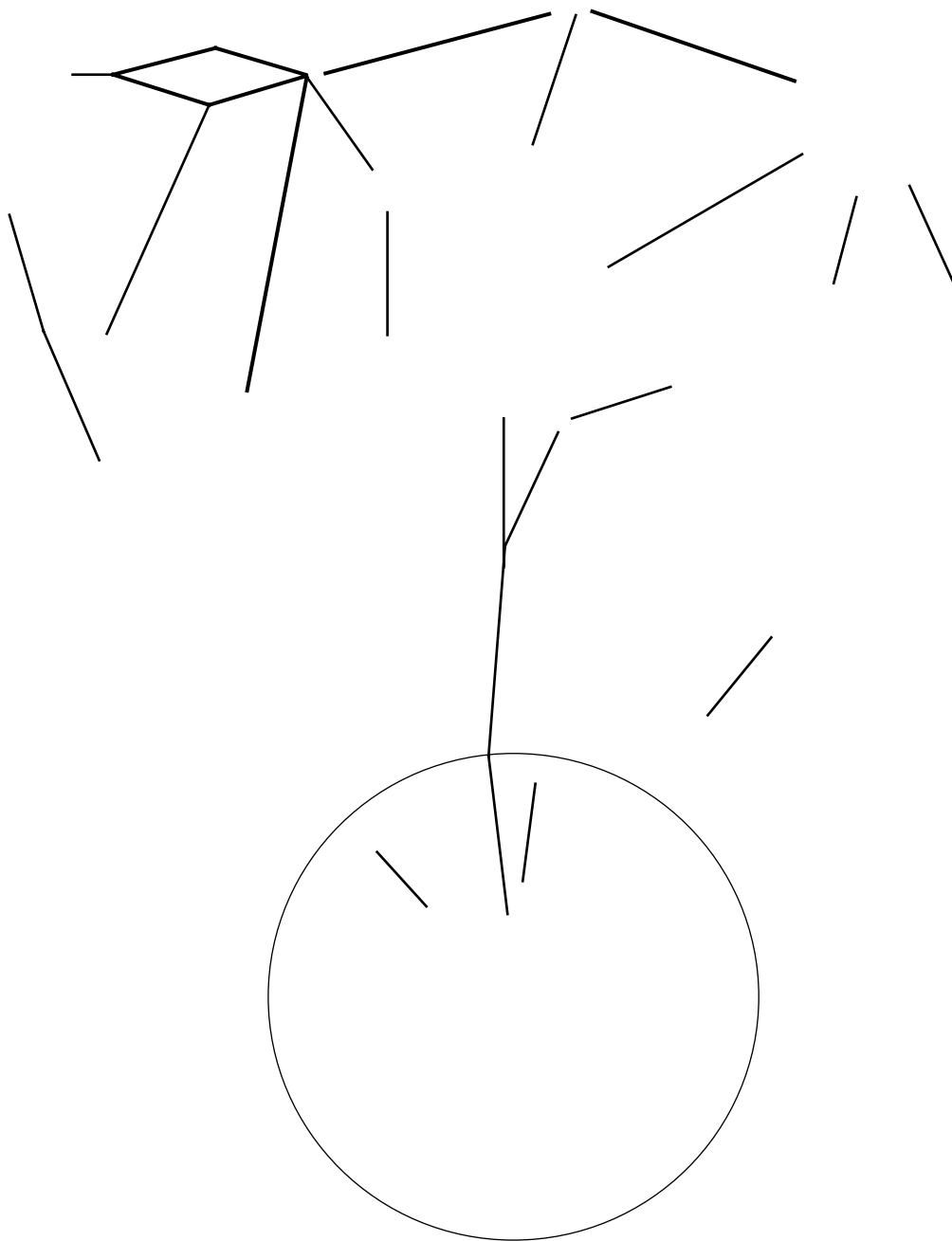
Cluster	HVSI motif
HV	
H	
V	16,298
K	16,224–16,311
U1	16,249
U2	16,051–(16,129C–16,189)
U3	16,343
U4	16,356
U5	16,270
U8	16,342
J	16,069–16,126
T	16,126–16,294
I	16,129–16,223–16,391
W	16,223–16,292
N1a	16,147A–16,223
N1b	16,145–16,176G–16,223–16,390
N9	16,257A–16,223–16,261
B	16,189
Y	16,126–16,231–16,266
F	16,304
A	16,223–16,290–16,319–16,362
M	16,223
C	16,223–16,298–16,327
Z	16,129–16,185–16,223–(16,224)–16,260–16,298
D	16,223–16,362
G	16,223–16,362

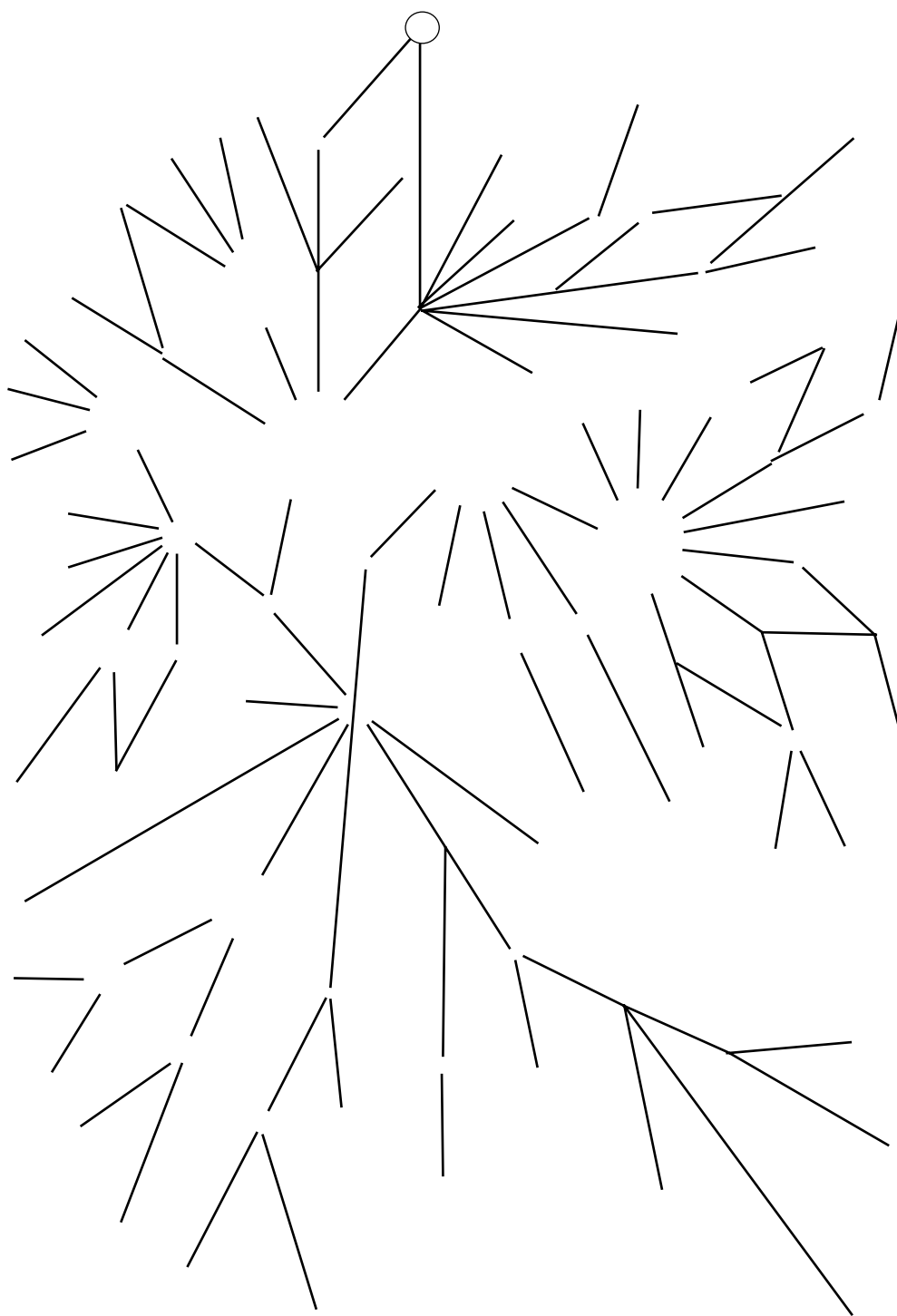
higher in the Finno-Ugric populations and maximal (24%) in Udmurts. The phylogenetic tree of this haplogroup was rather complex (Fig. 5), including several major variants (16,126–16,294, 16,126–16,294–16,296–16,304) and subcluster T1 (16,126–16,294–16,163–16,186–16,189).

Haplogroup J includes several founder types in Europe [24]. Variants of this haplogroup were observed at a frequency of 2–10% in all ethnic groups examined. The phylogenetic tree of this haplogroup is shown in Fig. 5.

Haplogroup I occurs predominantly in Northwestern Europe [5]. Its distribution in the populations of the Volga–Ural region was nonuniform. A single mtDNA variant of haplogroup I was relatively frequent (6%) in Mordvinians. A similar variant was observed in Bashkirs, Tatars, and Chuvash, but not in Mari, Udmurts, and Komi-Zyryans (Fig. 5).

The mitochondrial gene pool of the East Eurasian population includes haplogroups M, A, F, B, and Y.







Haplogroup M was well represented in all three ethnographic groups of Bashkirs. Its frequency was 28% on average, suggesting a substantial role of the Asian component for the Bashkir ethnogenesis. This is consistent with the early ethnic history of Bashkirs, since the ethnos is believed to originate from Turkic tribal confederations of the Altai, where ancient Bashkir and Mongolian tribes mixed together [12].

Haplogroup A occurs at the highest frequency in Siberian populations and rarely, if at all, in the indigenous populations of Southeastern Asia [2]. Thus the frequency of this haplogroup in Chukcha and Eskimo is 80%, higher than in all other world populations examined [25, 26]. We observed several variants of haplogroup A in Trans-Ural Bashkirs (6%); one of these was also found in Chuvash, Tatars, and Mari.

Haplogroups B and Y were detected only in Bashkirs. The former has been observed in Buryats, Tuvians, and Altaians of Southern Siberia, and occurs at a high (48%) frequency in Mongolia [28]. The common ancestor of these ethnic groups has been assumed to provide haplogroup B to native Americans. We observed the variants of this haplogroup at the eastern boundary of Europe (Table 2).

Haplogroup Y has first been identified in Eastern Asia [25]. This haplogroup is absent from most Siberian populations, and occur at a high frequency in the populations of Kamchatka, Sakhalin, and Korea. Presumably, haplogroup Y arose in ethnic groups of the Amur river basin [25] and was spread westwards by ethnic groups of the Altaian language family [22]. This family includes Bashkirs, who had one variant (16,126–16,231–16,266) of this haplogroup.

An interesting distribution of haplogroup F (6%) was observed in Bashkirs. Variants of this cluster (subgroup F1b) showed a high frequency and low diversity in Gaininsk Bashkirs (Fig. 6). Possibly, ethnogenesis of this group of Bashkirs was substantially affected by migrants from Central Asia, because similar mtDNA variants have earlier been observed in Kazakhs, Uigurs, and Mongols [22, 28]. In addition, our finding suggest a long-term isolation of this Bashkir group from its nearest neighbors, since haplogroup F was not found in any other ethnic group of the Volga–Ural region.

Cluster analysis showed that most mtDNA variants found in the ethnic groups of the Volga–Ural region represent haplogroups characteristic of Western Eurasia, testifying to a common origin of mtDNA lineages of West and East European populations. In our sample, the highest frequency of European mtDNA types was observed in Mordvinians, Mari, and Komi-Zyryans. On the other hand, the frequencies of the East Eurasian mtDNA variants were also high in the Volga–Ural region in contrast to Western and Central Europe. High frequencies of haplogroups G, D, C, Z,

and F in several Turkic (Bashkirs) and Finno-Ugric (Udmurts, Komi-Permyaks) ethnic groups suggest a substantial role of the Siberian and Central Asian components for the ethnogenesis of the Volga–Ural populations. Mordvinians, Komi-Zyryans, and Mari had low (3, 3, and 6%, respectively) frequencies of East Eurasian variants. Thus, the frequency and diversity of Siberian and Central Asian mtDNA types substantially increase in ethnic groups living at the boundary between Europe and Asia. It should also be noted that linguistically related populations genetically differ from each other.

REFERENCES

1. Cann, R., Stoneking, M., and Wilson, A., *Nature*, 1987, vol. 325, pp. 31–36.
2. Wallace, D.C., Brown, M.D., and Lott, M.T., *Gene*, 1999, vol. 238, pp. 211–230.
3. Denaro, M., Blanc, H., Johnson, M.J., *et al.*, *Proc. Natl. Acad. Sci. USA*, 1981, vol. 78, pp. 5768–5772.
4. Macaulay, V., Richards, M., Hickey, E., *et al.*, *Am. J. Hum. Genet.*, 1999, vol. 58, pp. 1309–1322.
5. Richards, M., Macaulay, V., and Bandelt, H.J., *Ann. Hum. Genet.*, 1998, vol. 62, pp. 241–260.
6. Torroni, A., Huoponen, K., Francalacci, P., *et al.*, *Genetics*, 1996, vol. 144, pp. 1835–1850.
7. Richards, M., Corte-Real, H., *et al.*, *Am. J. Hum. Genet.*, 1996, vol. 59, pp. 185–203.
8. Torroni, A., Lott, M., Cabell, M., *et al.*, *Am. J. Hum. Genet.*, 1994, vol. 55, pp. 760–776.
9. Torroni, A., Schurr, T., Yang, C., *et al.*, *Genetics*, 1992, vol. 130, pp. 153–162.
10. Torroni, A., Schurr, T., Cabell, M., *et al.*, *Am. J. Hum. Genet.*, 1993, vol. 53, pp. 563–590.
11. Cavalli-Sforza, L., Menozzi, P., and Piazza, A., *The History and Geography of Human Genes*, Princeton: Princeton Univ. Press, 1994.
12. Kuzeev, R.G., *Ann. Hum. Genet.* (Am. J. Hum. Genet.), 1999, vol. 58, pp. 993–1001.

20. Meinilä, M., Finillä, S., and Majamaa, K., *Hum. Hered.*, 2001, vol. 52, pp. 160–170.
21. Kivisild, T., Kaldma, K., Metspalu, M., *et al.*, in *Genomic Diversity*, Kluwer Academic, 1999, pp. 135–152.
22. Comas, D., Calafell, F., Mateu, E., *et al.*, *Am. J. Hum. Genet.*, 1998, vol. 63, pp. 1824–1838.
23. Metspalu, E., Kivisild, T., Kaldma, K., *et al.*, in *Genomic Diversity*, Kluwer Academic, 1999, pp. 121–134.
24. Finillä, S. and Majamaa, K., *J. Hum. Genet.*, 2001, vol. 46, pp. 64–69.
25. Schurr, T., Sukernik, R., Starikovskaya, Y., *et al.*, *J. Phys. Anthropol.*, 1999, vol. 108, pp. 1–39.
26. Starikovskaya, Y., Sukernik, R., Schurr, T., *et al.*, *Am. J. Hum. Genet.*, 1998, vol. 63, pp. 1473–1491.
27. VILLEMS, R., Adojaan, M., Kivisild, T., *et al.*, *The Roots of Peoples and Languages of Northern Eurasia*, vol. 1, Wiik, K. and Julku, K., Eds., Turku: Societas Historiae Fenno-Ugricae, 1998.
28. Kolman, C. and Sambuughin, N., *Genetics*, 1996, vol. 142, pp. 1321–1334.
29. Torroni, A., Sukernik, R.I., Schurr, T.G., *et al.*, *Am. J. Hum. Genet.*, 1993, vol. 53, pp. 591–608.
30. Bermisheva, M.A., Viktorova, T.V., and Khusnutdinova, E.K., *Genetika*, 2001, vol. 37, pp. 1002–1007.
31. Derenko, M., Malyarchuk, B., Dambueva, I., *et al.*, *Hum. Biol.*, 2000, vol. 72, pp. 945–973.