

Mitogenomic Diversity in Tatars from the Volga-Ural Region of Russia

Boris Malyarchuk,^{*,1} Miroslava Derenko,¹ Galina Denisova,¹ and Olga Kravtsova²

¹Genetics Laboratory, Institute of Biological Problems of the North, Far-East Branch of Russian Academy of Sciences, Magadan, Russia

²Department of Biochemistry, Kazan State University, Kazan, Russia

*Corresponding author: E-mail: malyarchuk@ibpn.ru.

Associate editor: Lisa Matisoo-Smith

Abstract

To investigate diversity of mitochondrial gene pool of Tatars inhabiting the territory of the middle Volga River basin, 197 individuals from two populations representing Kazan Tatars and Mishars were subjected for analysis of mitochondrial DNA (mtDNA) control region variation. In addition, 73 mitochondrial genomes of individuals from Mishar population were sequenced completely. It was found that mitochondrial gene pool of the Volga Tatars consists of two parts, but western Eurasian component prevails considerably (84% on average) over eastern Asian one (16%). Eastern Asian mtDNAs detected in Tatars belonged to a heterogeneous set of haplogroups (A, C, D, G, M7, M10, N9a, Y, and Z), although only haplogroups A and D were revealed simultaneously in both populations. Complete mtDNA variation study revealed that the age of western Eurasian haplogroups (such as U4, HV0a, and H) is less than 18,000 years, thus suggesting re-expansion of eastern Europeans soon after the Last Glacial Maximum.

Key words: mitochondrial DNA, genome sequencing, genetic diversity, Volga Tatars, Turkic-speaking populations.

Introduction

The Volga Tatars live in the central and eastern parts of European Russia and in western Siberia. They are the descendants of the Bulgar and Kipchak Turkic tribes who inhabited the western wing of the Mongol Empire, the area of the middle Volga River (Khalikov 1978; Kuzeev 1992). The Volga Bulgars settled on the Volga in the eighth century, where they mingled with Scythian- and Finno-Ugric-speaking peoples. After the Mongol invasion, much of the population survived and mixed with the Kipchak Tatars. Thus, in the Golden Horde time, the middle Volga River region became a melting pot of different people. In the 16th century, this area was conquered by “Ivan” the Terrible, first Tsar of Russia. Anthropologically, about 80% of the Volga Tatars belong today to Caucasoids and 20% to Mongoloids (Khalikov 1978). Linguistically, they speak language of a distinct branch of the Turkic group, within the Altaian family of languages.

According to results of studies of mitochondrial DNA (mtDNA) variation in populations of the Volga-Ural region, Tatars show intermediate frequencies of haplogroups characteristic of eastern Eurasia (about 12%) in relation to ethnic groups with highest (Finno-Ugric-speaking Udmurts and Komi-Permyaks and Turkic-speaking Bashkirs) and low (Finno-Ugric-speaking Mordvins, Mari, and Komi-Zyryans) frequencies of eastern Asian mtDNAs (Bermisheva et al. 2002). The presence of such mtDNAs in the mitochondrial gene pools of the Volga-Ural indigenous peoples suggests a substantial role of Siberian and central Asian populations in ethnic history of the Volga-Ural region. This region is also very important being a source of migration of eastern Europeans to the north of Europe. Genetic studies have shown that the Volga-Ural region may be a probable source for Saami and

Finnish mitochondrial diversity because of the presence of some mtDNA haplogroups characteristic to the Volga-Ural populations (U5b1b1, V, and Z1a) in Fennoscandia (Ingman and Gyllensten 2007). Meanwhile, molecular genetic data indicate multiple migrations from the east to the north of Europe, the first being 6.0–7.0 ka ago and at least one additional migration 2.0–3.0 ka ago (Ingman and Gyllensten 2007).

Although there are several examples of research on the Volga Tatar populations (Bermisheva et al. 2002; Orekhov 2002; Kravtsova 2006), only mtDNA hypervariable segment (HVS) I sequences and coding region restriction fragment length polymorphisms (RFLPs) have been analyzed in these populations, and there are no studies performed at the level of complete mitochondrial genome resolution. Meanwhile, further development of phylogeographic studies in Europe requires a considerable enlargement of databases of the complete mitochondrial genomes in large population samples. Here, we present an analysis of the complete mtDNAs of the Volga Tatars, with the purpose of studying their genetic diversity.

Materials and Methods

A population sample of 197 individuals from two areas of Republic of Tatarstan (Russian Federation), eastern ($n = 71$) (Aznakaevo) and western ($n = 126$) (Buinsk), was studied. Tatars from Aznakaevo belong to a group of Kazan Tatars, whereas Tatars from Buinsk belong to a group of Mishars. The Kazan Tatars and the Mishars are two major groups among the Volga Tatars, which are characterized by linguistic and ethnogenetic particularities (Kuzeev 1992). All individuals studied were maternally unrelated and originated from the area considered for this study. Appropriate

informed consent was obtained from all participants in this study.

DNA samples from the blood of individuals studied were used for mtDNA amplification and sequencing. Polymerase chain reaction amplification of mtDNA control region was performed using primers L15997 and H16547. The nucleotide sequences of mtDNA control region from nucleotide position (np) 15997–16526 in 197 Tatar individuals were determined on ABI 3130 Genetic Analyzer using BigDye v. 3.1 chemistry (Applied Biosystems, Foster City, CA). Complete sequences of mitochondrial genomes in 73 individuals were determined in the same way using the methodology described in detail by Torroni et al. (2001). DNA sequence data were analyzed using SeqScape v. 2.5 software (Applied Biosystems) and compared with the revised Cambridge reference sequence (Andrews et al. 1999). All samples sequenced for control region were subjected to RFLP analysis of coding region sites that were diagnostic of all major Eurasian clusters (haplogroups and subhaplogroups), on the basis of the hierarchical mtDNA RFLP scheme as described elsewhere (Malyarchuk et al. 2002; Derenko et al. 2007).

For comparative purposes, published mtDNA HVS I sequences and RFLP data in populations of the Volga-Ural region were used (Bermisheva et al. 2002; Orekhov 2002). In addition, complete or nearly complete mtDNA sequences pooled in PhyloTree (van Oven and Kayser 2009) were taken into analysis. For phylogeny construction, the length variation in the poly-C stretches at nps 16180–16193 and 309–315 and in the CA repeat at nps 514–524 was not used. The complete mtDNA trees were reconstructed by using median-network analysis with Network 4.5.1.0 (Bandelt et al. 1999) and mtPhyl program (<http://eltsov.org>), which is designed to reconstruct maximum parsimony phylogenetic trees. Both programs calculate haplogroup divergence estimates ρ and their error ranges as average number of substitutions in mtDNA clusters (haplogroups) from the ancestral sequence type (Saillard et al. 2000) using mutation rates based on 1) mtDNA complete genome variability data (one mutation every 3,624 years) and 2) synonymous substitutions (one mutation every 7,884 years) (Soares et al. 2009). To convert ρ values to age estimates with 95% confidence interval bounds, we used a calculator provided by Soares et al. (2009).

Haplotype and nucleotide diversity indices and their variances within populations were calculated using DnaSP version 5.0 (Librado and Rozas 2009). Neutrality tests by Elson et al. (2004) and Ruiz-Pesini et al. (2004) were performed using mtPhyl program (<http://eltsov.org>) to compare the ratio of the numbers of synonymous (S) and nonsynonymous (NS) substitutions in DNA sequences stratified into two classes on the basis of median-network analysis. The first class includes polymorphisms associated with haplogroups. The second class concerns private polymorphisms; these substitutions occur at the tips of individual branches within a network. The significance of the differences in NS:S ratios between two classes was determined on the basis of Fisher's exact test. All com-

Table 1. mtDNA Haplogroup Frequencies in the Volga Tatar Populations.

Haplogroup	Aznakaevo Tatars n = 71	Buinsk Tatars n = 126
A	4.2 (3)	3.2 (4)
C	5.6 (4)	0
D	9.9 (7)	4.8 (6)
G	0	1.6 (2)
M10	0	0.8 (1)
M7b	0	0.8 (1)
N9a	1.4 (1)	0
Y	0	0.8 (1)
Z	2.8 (2)	0
Eastern Asian component	23.9 (17)	11.9 (15)
H	32.4 (23)	34.1 (43)
I	5.6 (4)	1.6 (2)
J	7.0 (5)	7.1 (9)
K	0	4.8 (6)
M1	0	1.6 (2)
N1a,b,c	0	2.4 (3)
R2	1.4 (1)	0
T	2.8 (2)	4.8 (6)
T1	5.6 (4)	0.8 (1)
U1	0	0.8 (1)
U2	1.4 (1)	0
U3	0	1.6 (2)
U4	5.6 (4)	8.7 (11)
U5	2.8 (2)	4.0 (5)
U8a	0	4.0 (5)
HV0	2.8 (2)	6.3 (8)
W	7.0 (5)	5.6 (7)
X	1.4 (1)	0
Western Eurasian component	76.1 (54)	88.1 (111)

pletely sequenced mitochondrial genomes have been submitted to the GenBank database under accession numbers GU122975-GU123047 and EU567453-EU567455.

Results and Discussion

mtDNA Haplogroup Frequencies

In 197 individuals representing two populations of the Volga Tatars—the Kazan Tatars from Aznakaevo and the Mishars from Buinsk—we sequenced the mtDNA control region (between nps 15997 and 16526) (supplementary table S1, Supplementary Material online). Haplogroups were identified by means of RFLP analysis of the mtDNA-coding region. In general, 27 haplogroups were revealed in two populations studied (table 1). Most mtDNA haplotypes were assigned to western Eurasian haplogroups at frequencies of 76% in Aznakaevo population and 88% in Buinsk population. In Tatars, eastern Asian component seems to be heterogeneous, although only haplogroups A and D were revealed simultaneously in both populations. Western Eurasian haplogroups H, J, U4, and W were among the most common haplogroups in Tatar samples (table 1). Table 2 shows haplogroups frequency distribution in Tatars in comparison with neighboring eastern European populations, such as Bashkirs, Chuvash, Mari, Mordvins, Udmurts, Karelians, and Russians. Some

Table 2. Haplogroup Distributions (%) in Tatars in Comparison with Other Eastern European Populations.

Haplogroup	Tatars ^a	Tatars ^b	Bashkirs ^b	Chuvash ^b	Mari ^b	Mordvins ^b	Udmurts ^b	Karelians ^c	Russians ^{d-f}
H	33.5	30.7	12.2	27.3	40.4	42.2	21.8	46.9	41.6
HV0	5.1	3.9	3.2	7.3	11.0	4.9	0	6.5	5.0
HV*	0	0.9	0.5	0	1.5	1.0	0	0	3.2
U1	0.5	0.9	0	0	0	0	0	2.1	0.3
U2	0.5	0.9	0.5	0	0	6.9	9.9	0.4	1.2
U3	1.0	2.2	0	1.8	0	0	0	0.6	1.2
U4	7.6	7.0	12.7	16.4	10.3	2.0	4.0	2.7	3.3
U5	3.6	10.5	13.6	14.5	14.0	15.7	8.9	17.0	10.9
U7	0	0	0	0	0	0	0	0	0.5
U8	2.5	0	0.5	1.8	0	0	0	2.7	0.8
K	3.0	5.7	1.4	7.3	2.2	0	0	1.6	2.9
U*	0	2.2	0	1.8	0	2.0	2.0	1.4	0
J	7.1	7.5	3.2	5.5	7.4	7.8	2.0	4.5	8.4
T*	4.1	6.6	1.4	0	3.7	5.9	8.9	2.0	7.2
T1	2.5	2.6	4.1	3.6	1.5	2.0	14.9	1.8	2.4
R*	0.5	0.4	0	0	0	1.0	6.9	0.8	0.7
I	3.0	0.9	1.4	1.8	0.7	5.9	0	2.2	2.4
W	6.1	1.8	0.5	0	0	0	0	1.8	1.7
X	0.5	0	0	0	0	0	0	0	1.6
N1a	0.5	0.4	3.6	1.8	0	0	0	0	0.7
N1b	0.5	1.8	0	0	0	0	0	0	0.7
N1c	0.5	0	0	0	0	0	0	0	0
M*	1.0	2.2	1.4	1.8	0.7	0	0	0	0.2
C	2.0	1.8	11.8	1.8	0.7	2.0	3.0	0	0.3
Z	1.0	0.4	0.9	0	2.9	0	5.0	0.4	0.2
D	6.6	2.6	9.0	3.6	1.5	1.0	11.9	3.7	0.6
G	1.0	1.8	4.5	0	0	0	0	0.2	0.2
M1	1.0	0	0	0	0	0	0	0	0.3
A	3.6	3.1	3.6	1.8	1.5	0	1.0	0	0.1
B	0	0	0.9	0	0	0	0	0	0
F	0	0	6.3	0	0	0	0	0	0
N9a	0.5	0.9	1.4	0	0	0	0	0	0.1
Y	0.5	0	0.5	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0.2
Others	0	0.4	1.4	0	0	0	0	0.7	0.2
Sample size	197	228	221	55	136	102	101	512	873

^a Data for populations are from present study.

^b Data for populations are from Bermisheva et al. (2002).

^c Data for populations are from Lappalainen et al. (2008).

^d Data for populations are from Malyarchuk et al. (2002).

^e Data for populations are from Malyarchuk et al. (2004).

^f Data for populations are from Grzybowski et al. (2007).

differences between Tatar samples investigated in different studies can be revealed—for instance, the lower frequency of haplogroup U5 or higher frequency of haplogroups W and D in our sample ($P < 0.05$, t -test). However, frequency of eastern Asian mtDNAs was similar in both samples (16.2% and 12.8% in our and Bermisheva et al.'s [2002] study, respectively), thus placing Tatars between the Volga-Ural region populations with high (>20% in Bashkirs and Udmurts) and moderate (<10% in Chuvash, Mari, and Mordvins) frequencies of eastern Asian mtDNA component (table 2).

The pairwise nucleotide difference distributions for control region sequences in Buinsk Tatars are clearly bell-shaped and unimodal being consistent with exponentially growing populations (Rogers and Harpending 1992). Average numbers of nucleotide differences have similar means of 4.496 and 4.781 in Buinsk and Aznakaevo populations, respectively. These estimations are in accordance with previous data for European populations (Comas et al.

1997). Comparison of two Tatar populations studied indicates that only Buinsk Tatars yielded significantly negative values for both Tajima's D ($D = -2.164$, $P < 0.01$) and Fu and Li's F^* ($F^* = -3.312$, $P < 0.02$) neutrality tests, thereby pointing to the signature of population expansion. Meanwhile, low between-population differences were found by means of G_{ST} and F_{ST} testing ($G_{ST} = 0.39\%$, $F_{ST} = 0.46\%$).

Complete mtDNA Sequence Variation in Tatars

Complete mtDNA sequences were determined in 73 maternally unrelated individuals from Buinsk population. These individuals were randomly selected from Buinsk sample with known mtDNA control region sequences. However, in our sampling strategy, we wished to characterize the whole diversity of nonidentical mtDNA control region haplotypes, so that strategy could be slightly biased. We found that 73 mtDNA sequences representing Tatar gene pool contained 507 polymorphic (segregating) sites.

In the protein-coding region, 335 nps were polymorphic, 113 nucleotide substitutions were nonsynonymous, and 249 substitutions were synonymous. In RNA-coding genes, 22 positions were polymorphic in transfer RNA genes, and 15 and 23 positions were polymorphic in 12S and 16S ribosomal RNA genes, respectively. The gene diversity of mtDNA sequences in Tatars is 0.998, and the average number of nucleotide differences is 35.197, thus demonstrating high genetic variability. Results of all tests on neutrality were significant (Tajima's *D* value is -2.333 , $P < 0.01$; Fu and Li's *D* and *F* values are -4.679 and -4.45 , respectively, $P < 0.02$; and Fu's *F* is -31.222 , $P < 0.0001$), thus showing consequences of population expansions (Excoffier and Schneider 1999).

To assess the role of selection, the numbers of synonymous and nonsynonymous substitutions in the 13 protein-coding genes were investigated using neutrality tests described by Elson et al. (2004) and Ruiz-Pesini et al. (2004). Results of analysis of neutrality indices do not support a model in which selection may be a major force influencing on mtDNA variability ($IT = 1.1$, $Ni = 0.91$, $P > 0.05$) (supplementary table S2, Supplementary Material online). However, results of gene-by-gene analysis of synonymous and nonsynonymous substitutions suggest the possible operation of positive selection on the *ND3* gene ($Ni = 0.08$, $P = 0.043$), but the number of nucleotide substitutions was relatively low. Thus, further analysis with much larger sequence sets of eastern European origin is required to confirm this result.

Topology of mtDNA Phylogenetic Networks in Tatars

The median network of 73 complete mtDNA sequences is shown in supplementary figure S1 (Supplementary Material online). This tree was constructed based on the existing classification of mtDNA haplogroups (van Oven and Kayser 2009). As seen in the tree, the majority of mtDNAs belong to a singular haplotypes within already known subhaplogroups. Specific subclusters of mtDNA haplotypes were revealed only in several cases—for W3 (two haplotypes), V1a (two), T2b* (three), U4b1b (two), U4d1 (two), and U4d as a whole (four).

However, some of the singular haplotypes appear to be informative for further development of mtDNA classification. Sample 23_Tm could be assigned to A10 according to nomenclature suggested by van Oven and Kayser (2009). However, phylogenetic analysis of complete mtDNAs (fig. 1) reveals that this sample belongs to haplogroup A8, which is defined now by transition at np 64 and consists of two related groups of lineages—A8a, with control region motif 146-16242 (previously defined as A8 by Derenko et al. [2007]), and A8b, with motif 16227C-16230 (supplementary table S3, Supplementary Material online). Analysis of HVS I and II sequences in populations indicates that transition at np 64 appears to be a reliable marker of haplogroup A8 (supplementary table S3, Supplementary Material online). The only exception, the probable back mutations at nps 64 and 146, has been described in Koryak haplotype

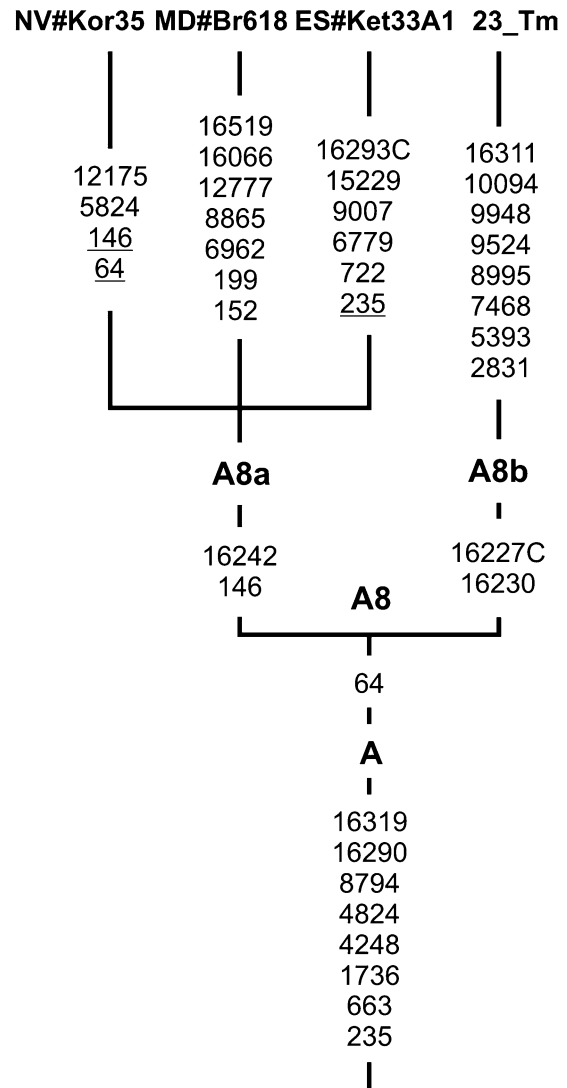


Fig. 1. Complete mtDNA sequence-based phylogenetic tree of haplogroup A8. The tree is rooted in haplogroup A. Numbers along links refer to substitutions scored relative to the revised Cambridge reference sequence (Andrews et al. 1999). Transversions are further specified, and back mutations are underlined. Three additional complete sequences were taken from the literature (Starikovskaya et al. 2005; Derenko et al. 2007; Volodko et al. 2008) and designated by ES, MD, and NV, respectively, followed by “#” and the original sample code.

EU482363 by Volodko et al. (2008). Therefore, parallel transversions at np 64 define not only Native American clusters of haplogroup A2, that is, its node A2c'd'e'f'g'h'i'j'k'n'p (Achilli et al. 2008; van Oven and Kayser 2009), but also northern Eurasian haplogroup A8. Both A8 and subhaplogroups are spread at relatively low frequencies in populations of central and western Siberia and in the Volga-Ural region. A8a is present even in Transylvania at frequency of 1.1% among Romanians, thus indicating that the presence of such mtDNA lineages in Europe may be mostly a consequence of medieval migrations of nomadic tribes from Siberia and the Volga-Ural region to Central Europe (Malyarchuk et al. 2006; Malyarchuk, Derenko, et al. 2008).

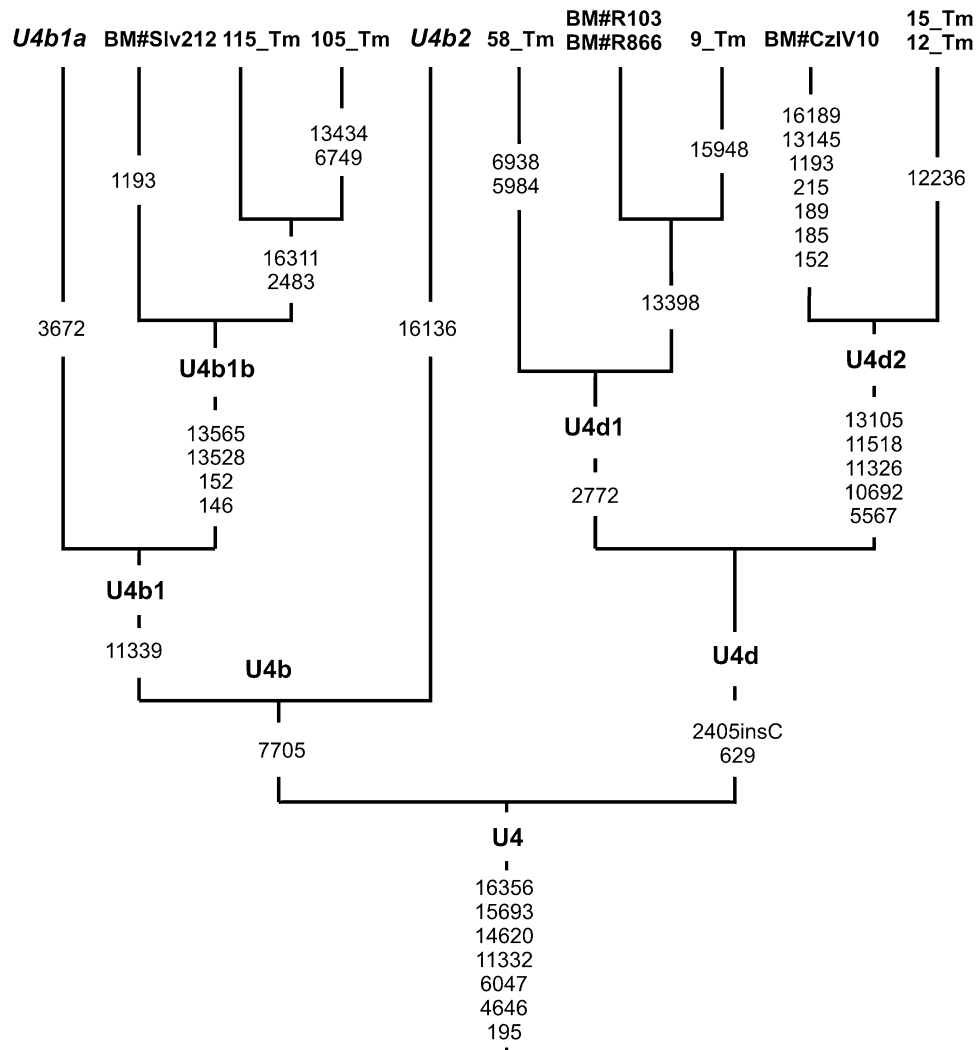


Fig. 2. Complete mtDNA sequence-based phylogenetic tree of subhaplogroups U4b and U4d. The tree is rooted in haplogroup U4. Numbers along links refer to substitutions scored relative to the revised Cambridge reference sequence (Andrews et al. 1999). Additional complete sequences were taken from the literature (Malyarchuk, Grzybowski, et al. 2008) and designated by BM followed by “#” and the original sample code. For subhaplogroups U4b1a and U4b2, only diagnostic mutations are shown.

Another case requiring further phylogenetic specification is haplogroup N1c. We have sequenced N1c haplotype of individual 56_Tm and found that transition at np 11914 should be added to haplogroup-specific motif indicated in PhyloTree (van Oven and Kayser 2009). This is because two individuals belonging to two different branches of N1c share this mutation (supplementary fig. S2, Supplementary Material online). In addition, in haplogroups Y1b and T2f1, we have found that Tatar mtDNAs had shortcut haplogroup motifs, with a lack of transitions at nps 15221 for Y1b and 15028 for T2f. Taking into account that 15221 is a fairly conserved nonsynonymous mutation and 15028 is even more rare (with zero occurrences in 2,196 complete mitochondrial genomes surveyed by Soares et al. [2009]), one can assume that their haplogroup-specific motifs should be shortened.

Haplogroup U4 is one of the most frequent in populations of the Volga-Ural region and western Siberia (Bermisheva et al. 2002; Derbeneva, Starikovskaya, Volodko, et al. 2002;

Derbeneva, Starikovskaya, Wallace, et al. 2002; Naumova et al. 2008). In Tatars, we sequenced ten mitochondrial genomes and found that they fall into three subhaplogroups—U4a, U4b, and U4d. Four sequences belong to U4a1*, U4a2b, and U4a2c1 subclusters (supplementary fig. S1, Supplementary Material online). Figure 2 demonstrates that haplotypes of two Tatar individuals 105_Tm and 115_Tm are clustered together with Slovak haplotype from our previous study (Malyarchuk, Grzybowski, et al. 2008) into subhaplogroup U4b1b. Haplogroup U4d could be subdivided into two subhaplogroups: U4d1 characterized by transition at np 2772 and U4d2 defined by motif 5567-10692-11326-11518-13105. Haplotypes of two Tatar individuals 58_Tm and 9_Tm are clustered with two Russians into subcluster U4d1, whereas Tatar (12_Tm and 15_Tm) and Czech haplotypes are combined into subcluster U4d2 (fig. 2).

Using the complete sequence and the synonymous mtDNA clocks (Soares et al. 2009), the coalescence age

Table 3. Age Estimates of mtDNA Haplogroups U4, H, HV0a, and V with ρ and from Two Different Molecular Clocks.

Haplogroup	Sample Size	ρ Estimates (95% confidence interval), in ka ago	
		Complete Genome	Synonymous Positions
U4	61	16.5 (10.7 to 22.4)	17.3 (5.6 to 29.1)
U4a	44	13.1 (7.8 to 18.5)	6.8 (4.0 to 9.5)
U4a1	14	9.6 (5.3 to 14.1)	11.3 (4.3 to 18.2)
U4a2	29	7.8 (4.7 to 11.0)	4.1 (1.5 to 6.7)
U4b	6	16.2 (7.8 to 24.9)	21.0 (5.1 to 36.9)
U4d	7	12.6 (5.6 to 19.8)	15.7 (2.2 to 29.3)
HV0a	40	14.4 (7.0 to 22.0)	14.6 (−0.1 to 30.3)
V	39	11.8 (6.5 to 17.2)	6.9 (2.1 to 11.6)
V1a	17	7.9 (2.5 to 13.5)	2.8 (−0.3 to 5.8)
V1a1a1	6	0.85 (−0.4 to 2.1)	0
V3a	6	4.8 (0.02 to 9.7)	7.9 (−0.5 to 21.0)
H	16	17.9 (13.6 to 22.3)	16.2 (9.4 to 23.0)

for 61 mitochondrial genomes of haplogroup U4 (from the present and our previous study) is ~ 17.0 ka ago (table 3). Coalescence time estimates for subhaplogroups vary from 13.0 to 16.0 ka ago for mutation rate based on complete mtDNA variability and from 7.0 to 21.0 ka ago for synonymous mutation rate. Neutrality testing demonstrates that selection does not influence on variability of haplogroup U4 in eastern Europe ($N_i = 0.46$, $P > 0.1$).

Previous studies have shown that the Volga-Ural region has an important role in the peopling of the northeast of Europe. It has been found that some Saami and Finnish mtDNA lineages (such as haplogroups U5b1b1, V, and Z1a) shared a common ancestor with lineages from the Volga-Ural region (Tambets et al. 2004; Achilli et al. 2005; Ingman and Gyllensten 2007). The divergence time for the Saami and Finnish haplogroup V sequences was estimated as 7.6 ka ago and for U5b1b1 as 5.5–6.6 ka ago (Ingman and Gyllensten 2007). Much lower age of haplogroup Z1a (2.7 ka ago) suggests also a more recent contribution of people from the Volga-Ural region to the Saami gene pool. In the present study, we sequenced five mitochondrial genomes of Tatars belonging to haplogroup HV0, so these additional data appear to be useful for molecular dating of haplogroup V. We analyzed 32 published haplogroup V mitochondrial genomes from populations of Finns and Saami (Finnilä et al. 2001; Ingman and Gyllensten 2007) and three novel genomes from populations of Russians and Czechs (EU567453–EU567455) (supplementary fig. S3, Supplementary Material online). As a result, we have found that three mtDNAs of Tatars fall into subcluster V1a that is very frequent among Finns (Finnilä et al. 2001). It is noteworthy that Tatars and Finns share mtDNAs from this subcluster on a large time span—from 7.9 and 2.8 ka ago (for complete genome and synonymous rates, respectively) for subcluster V1a to 0.85 ka ago and even less (for complete genome and synonymous rates) for small subcluster V1a1a1 (table 3). In addition, analysis of subcluster V3a demonstrates that divergence between Russian/Finnish mtDNAs is estimated as 4.8 and 7.9 ka ago (for complete genome and synonymous rates, respectively).

Based on complete mtDNA variation of eastern Europeans (for 40 mitochondrial genomes from populations of

Finns, Saami, Tatars, Russians, and Czechs), haplogroup HV0a dates to ~ 14.5 ka ago, fitting the time of expansion from European glacial refuge zones (the Franco-Cantabrian, Balkan, and Ukrainian ones) (table 3). Meanwhile, haplogroup V (39 genomes of eastern Europeans) dates to 11.8 ka ago for complete genome rate and to 6.9 ka ago for synonymous rate, that is, somewhat less than previously reported dating results for haplogroup V in Europe—13.7 (12.1–15.2) ka ago for complete genome rate and 12.2 (10.0–14.3) ka ago for synonymous rate (Soares et al. 2009).

Conclusions

In summation, mitochondrial gene pool of the Volga Tatars is characterized by high level of diversity and it can be seen as composite of both western and eastern Eurasian mtDNA haplogroups. In the Volga Tatars, there are no specific mtDNA clusters of eastern Asian origin, allowing us to suggest early presence of eastern Asians/Siberians in the Volga River basin. On the contrary, western Eurasian mtDNA lineages appear to be very diverse. The age of haplogroups U4 and HV0a, which is less than 18.0 ka ago, suggests re-expansion of eastern Europeans soon after the Last Glacial Maximum, as does the age of haplogroup H, which dates to ~ 17.0 ka ago based on mtDNA diversity in Tatars (table 3). Much more detailed information on mtDNA diversification process in eastern Europe will be obtained after considerable enlargement of database of complete mitochondrial genomes created in the present study.

Supplementary Material

Supplementary tables S1–S3 and figures S1–S3 are available at the *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We are grateful to Nikolay Eltsov for help in this study. The study was supported by grants from the Program of Precedium of Russian Academy of Sciences “Biodiversity”

(09-I-P23-10) and the Russian Foundation for Basic Research (07-04-00445).

References

- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One*. 3:e1764.
- Achilli A, Rengo C, Battaglia V, et al. (13 co-authors). 2005. Saami and Berbers—an unexpected mitochondrial DNA link. *Am J Hum Genet*. 76:883–886.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 23:147.
- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16:37–48.
- Bermisheva M, Tambets K, VILLEMS R, Khusnutdinova E. 2002. Diversity of mitochondrial DNA haplogroups in ethnic populations of the Volga-Ural region. *Mol Biol (Mosk)*. 36:990–1001.
- 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.
- Derbeneva OA, Starikovskaya EB, Volodko NV, Wallace DC, Sukernik RI. 2002. Mitochondrial DNA variation in Kets and Nganasans and the early peoples of Northern Eurasia. *Genetika* 38:1554–1560.
- Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI. 2002. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 70:1009–1014.
- Derenko M, Malyarchuk B, Grzybowski T, et al. (12 co-authors). 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet*. 81:1025–1041.
- Elson JL, Turnbull DM, Howell N. 2004. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am J Hum Genet*. 74:229–238.
- Excoffier L, Schneider S. 1999. Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. *Proc Natl Acad Sci U S A*. 96:10597–10602.
- Finnilä S, Lehtonen MS, Majamaa K. 2001. Phylogenetic network for European mtDNA. *Am J Hum Genet*. 68:1475–1484.
- Grzybowski T, Malyarchuk BA, Derenko MV, Perkova MA, Bednarek J, Woźniak M. 2007. Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *Forensic Sci Int Genet*. 1:141–147.
- Ingman M, Gyllensten U. 2007. A recent genetic link between Sami and the Volga-Ural region of Russia. *Eur J Hum Genet*. 15:115–120.
- Khalikov AK. 1978. Origin of the Volga-Ural Tatars. Kazan (Russia): Kazan Book Publishers. Russian.
- Kravtsova OA. 2006. Molecular-genetic analysis of ancient and modern DNA samples [PhD thesis]. [Kazan (Russia)]: Kazan State University. Russian.
- Kuzeev RG. 1992. Ethnic groups of the central Volga region and southern Urals. Moscow (Russia): Nauka. Russian.
- Lappalainen T, Laitinen V, Salmela E, Andersen P, Huoponen K, Savontaus ML, Lahermo P. 2008. Migration waves to the Baltic Sea region. *Ann Hum Genet*. 72:337–348.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Malyarchuk B, Derenko M, Grzybowski T, Lunkina A, Czarny J, Rychkov S, Morozova I, Denisova G, Miścicka-Śliwka D. 2004. Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. *Hum Biol*. 76:877–900.
- Malyarchuk B, Derenko M, Perkova M, Vanecek T. 2008. Mitochondrial haplogroup U2d phylogeny and distribution. *Hum Biol*. 80:565–571.
- Malyarchuk B, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miścicka-Śliwka D. 2002. Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet*. 66:261–283.
- Malyarchuk B, Grzybowski T, Derenko M, Perkova M, Vanecek T, Lazur J, Gomolcak P, Tsybovsky I. 2008. Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Mol Biol Evol*. 25:1651–1658.
- Malyarchuk BA, Vanecek T, Perkova MA, Derenko MV, Sip M. 2006. Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. *Hum Biol*. 78:681–696.
- Naumova OY, Rychkov SY, Morozova IY, Khaiat SS, Semikov AV, Zhukova OV. 2008. Mitochondrial DNA diversity in Siberian Tatars of the Tobol-Irtysh basin. *Genetika* 44:257–268.
- Orekhov VA. 2002. Characteristics of mtDNA types in three ethnic groups of the European part of Russia. [PhD thesis]. [Moscow (Russia)]: Vavilov Institute of General Genetics. Russian.
- Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol*. 9:552–569.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303:223–226.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Norby S. 2000. MtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet*. 67:718–726.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*. 84:740–759.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, et al. (11 co-authors). 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet*. 69:67–89.
- Tambets K, Rootsi S, Kivisild T, et al. (44 co-authors). 2004. The western and eastern roots of the Saami—the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet*. 74:661–682.
- Torroni A, Rengo C, Guida V, et al. (11 co-authors). 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet*. 69:1348–1356.
- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 30:E386–E394.
- Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV, Wallace DC, Sukernik RI. 2008. Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocene peopling of the Americas. *Am J Hum Genet*. 82:1084–1100.