
HUMAN GENETICS

Differentiation of the Mitochondrial Subhaplogroup U4 in the Populations of Eastern Europe, Ural, and Western Siberia: Implication to the Genetic History of the Uralic Populations

B. A. Malyarchuk

*Institute of the Biological Problems of the North, Russian Academy of Sciences, Magadan, 685000 Russia;
fax: 41322-34463; e-mail: malyar@ibpn.kolyma.ru*

Received September 9, 2003

Abstract—Phylogenetic relationships between the sequences of the mitochondrial DNA (mtDNA) hypervariable segment 1, belonging to subhaplogroup U4, were examined in the populations of Eastern Europe, Ural, and Northwest Siberia. It was shown that the frequency of subhaplogroup U4, as well as its proportion in the U-component of the gene pools, increased eastwards, reaching maximum values in the populations of Northwest Siberia. Phylogenetic analysis it was showed that the appearance of specific U4-lineage (16113C–16356–16362) in the ancestors of Mansi was most likely caused by its divergence from the East European cluster 16356–16362 in the Late Upper Paleolithic (18566 ± 12915 years before present). Other U4 mtDNA lineages (16189–16356 and 16311–16356), typical mostly of the indigenous populations of Northwest Siberia (Mansi, Nganasans, and Kets) may have formed during the Neolithic–early Bronze Age (6055 ± 3599 years before present, on average). It seems likely that the isolation of ancient populations inhabiting the region between the Ob' and Yenisei rivers was the key factor, providing the appearance of the unique Caucasoid mtDNA lineages in their gene pools. These results were consistent with the traditional point of view on the mixed origin of the Finno-Ugric populations of the Volga–Ural region and West Siberia, resulted from the genetic relationships between the populations of Europe and Asia.

INTRODUCTION

One of the first phylogenetic analysis of the mitochondrial DNA (mtDNA) variation in the populations of Western Europe showed that the most ancient in Europe mitochondrial haplogroup U (of the age that exceeds 50 000 years) was represented by a number of subhaplogroups differing by the time and place of their origin [1]. One of these was subhaplogroup U4, with an evolutionary age of more than 25 000 years [2]. Based on the level and time of the U4 mtDNA types divergence in the populations of Near East and Europe ($25\,650 \pm 9350$ and $20\,400 \pm 4300$ years, respectively), the Near Eastern (Anatolian) origin of this mtDNA subhaplogroup is suggested. The highest prevalence of the subhaplogroup, however, was detected in Europe and in particular in its northwestern part [2, 3]. The study of the geographic frequency distribution of subhaplogroup U4 in the European populations showed that its highest frequency (5.5%) was typical of Baltic–Finish populations of Northeastern Europe (Finns, Karelians, and Estonians), while in the populations of Central and Northwestern Europe, the frequency of U4 did not exceed 3% [3, 4].

The study of the populations of Eastern Europe showed that subhaplogroup U4 belonged to the mitochondrial lineages mostly prevalent among the populations of the Volga–Ural region [5]. The frequency of this subhaplogroup among the Finno–Ugric and Turkic

populations from the Volga and Cis-Ural regions varied from 2 to more than 20%, and the highest frequencies were observed among Komi-Zyryans, Chuvashes, and Bashkirs (24.2, 16.4, and 12.7%, respectively) [5]. Analysis of the more eastern, Trans-Ural and Siberian, populations showed that subhaplogroup U4 was widely distributed (with the frequency of 16 to 29%) among the indigenous populations of Northwest Siberia, Mansi, Nganasans, and Kets [6, 7]. This observation served as a basis for the suggestion that the high frequency of the U4 mtDNA types in the gene pools of the populations from the Eastern Cis-Ural and the adjacent Siberian regions can be considered as a genetic trace of the Upper Paleolithic proto-Eurasian population (“an independent Eurasian structure,” according to Bunak [8]), currently preserved only to the east of the Ural Mountains, but initially formed in the Middle East [6, 7]. The study of the indigenous populations of South Siberia also pointed to the presence of the U4 mtDNA types in the gene pools of some Altai–Sayan populations. In particular, subhaplogroup U4 with the frequency of about 6% was found in Altai-Kizhi and Khakassians [9]. Thus, reviewing the data on the prevalence of the U4 mtDNA types in the populations of Europe and the adjacent territories of Ural and West Siberia, it becomes evident that the frequency of this subhaplogroup increases eastwards, reaching maximum in the populations inhabiting the territories between the Ob' and Yenisei rivers. However, the pattern of phylogenetic relationships

between the U4 mtDNA types present in different East European and Siberian populations remains unclear. In this study, the diversity of subhaplogroup U4 and its individual clusters in different populations of Eastern Europe, Ural, and Siberia was examined. In our study, we used a phylogenetic approach that considers both the frequency and diversity of monophyletic DNA clusters, thus providing the most adequate conclusions on the evolutionary history of the gene pools and their individual components [1, 2].

MATERIALS AND METHODS

The phylogenetic analysis was based on the data on mtDNA polymorphism in western Finnish populations of the Baltic region ($n = 373$), represented in the studies [10, 11] by Finns ($n = 242$), Estonians ($n = 48$), and Karelians ($n = 83$); in eastern Finnish and Turkic populations of the Volga–Ural region ($n = 979$), represented in the study [5] by Mordovians ($n = 102$), Maris ($n = 136$), Komi-Zyryans ($n = 62$), Komi-Permyaks ($n = 74$), Tatars ($n = 228$), Chuvashes ($n = 55$), Udmurts ($n = 101$), and Bashkirs ($n = 221$); in the populations of Northwest Siberia ($n = 218$), represented by Nentsi ($n = 58$), Mansi ($n = 98$), Nganasans ($n = 24$), and Kets ($n = 38$) [6, 7, 12]; in the populations of South Siberia ($n = 480$), represented in the study [13] by Altaians ($n = 110$), Khakasians ($n = 53$), Buryats ($n = 91$), Sojots ($n = 30$), Tuvians ($n = 90$), Todjins ($n = 48$), and Tofalars ($n = 58$); in Russian populations of Stavropol krai, Krasnodar krai, Orel oblast, Belgorod oblast, Saratov oblast, and Nizhni Novgorod oblast ($n = 391$, elicited from the study [14]). The summarized sample size constituted 2441 individuals.

In almost all samples, the attribution of the mtDNA HVS1 types to subhaplogroup U4 was determined based on the data on the distribution of the markers, distinguishing haplogroup U (variant +12308HinfI, mutation 12308AG) and its subhaplogroup U4 (variant +4643RsaI, mutation 4646TC), presented in the studies mentioned above. The exceptions were the samples of Finns, Estonians, and Karelians (data taken from [10]), for which the data on the prevalence of the U4 mtDNA HVS1 types were presented in accordance with the database described in [2].

Evolutionary relationships between the U4 mtDNA HVS1 types were examined using RM (reduced-median) algorithm of the median network method [15] as implemented in the Network 3.1.1.1 software package (www.fluxus-engineering.com). Similarly to the other maximum parsimony methods, the median network method is based on the principle of minimization of the number of evolutionary events, leading to the DNA changes. The results of the analysis are presented graphically as phylogenetic nets, combining several plausible dendrogram topologies. Topology combination results in the appearance of n -dimensional cycles, the number of which can be reduced by decreasing the weight of the hot-spot nucleotide positions. Analyzing

U4 sequences of mtDNA, we reduced the hot-spot weights at positions 16093, 16129, 16189, and 16362 (according to [16]) to 1/10 relative to the other HVS1 positions. The hot-spot weight reduction leads to more fanlike branching of the mtDNA types within the clusters, providing far more precise evaluation of the nucleotide sequences divergence levels [2, 17].

Genetic distances ρ between the mtDNA sequences were calculated as the mean number of mutations between the founder genes and the derivative mtDNA types belonging to the corresponding monophyletic clusters [18]. Determination of the evolutionary ages of the mtDNA clusters was based on the concept that for the HVS1, genetic distance $\rho = 1$ corresponds to the time interval of 20180 years [18]. This rate is close to the other mtDNA HVS1 mutation rates suggested earlier [19, 20]. Standard deviations for ρ were calculated according to [17] using Network 3.1.1.1 software program. It should be noted in this respect that statistical robustness of the timing data based on the mtDNA HVS1 sequence variation is doubtful, since there is still no agreement on the initial assumptions underlying the method of the mutation rate estimation (see, for example, [19]). In addition, some discrepancies between the mutation rate values determined based on the family (pedigree) and population data were observed [21]. Nevertheless, in the present study, similarly to most of the other phylogenetic works (including those, used here for comparison [2, 4, 5]), the mtDNA HVS1 mutation rate mentioned above was used.

RESULTS AND DISCUSSION

Analysis of haplogroup U4 frequency distribution patterns in the populations of Europe, Ural, and Siberia (table) showed that in Russian populations of Eastern Europe, the frequency of U4 is comparable to that observed in the populations of Central and Northwestern Europe (about 3% [3]). Among the Finno–Ugric populations the frequency of U4 gradually increases from 5.9% in Baltic–Finnish populations (Finns, Estonians, and Karelians) to 10% in Eastern-Finnish (and also Turkic) populations of the Volga–Ural region, reaching maximum values (17%, on average) in Ugro–Samoyedic populations of Western Siberia. Among the South Siberian populations examined (Altaians, Khakasians, Buryats, Sojots, Tuvians, Todjins, and Tofalars), subhaplogroup U4 was detected only in Altai-Kizhi (5.5%) and Khakasians (7%) [13]. Note that similar changes were observed for the proportion of the U4 types among all U mtDNA types in the populations tested. Specifically, in the populations of Northwestern and Central Europe, and also in Russians, the proportion of U4 within haplogroup U constitutes 13 to 17%, while in the indigenous populations of Eastern Europe, Ural, and West Siberia it gradually increases, reaching maximum value (67%) in the populations of Northwest Siberia (table). To the south of Siberia, the frequency of U4 decreases, although its