Mitochondrial DNA sequence diversity in Russians

Vladimir Orekhov[a,*], Andrey Poltoraus[b], Lev A. Zhivotovsky[a], Viktor Spitsyn[c], Pavel Ivanov[b], Nikolay Yankovsky[a]

[a] *Vavilov Institute of General Genetics, Moscow 117809, Russia*
[b] *Engelhardt Institute of Molecular Biology, Moscow 117894, Russia*
[c] *Medical-Genetic Scientific Centre, Moscow 115478, Russia*

Received 13 January 1999

Abstract The article presents the results of the first regular study of Russian populations by sequencing the control region of mitochondrial DNA (mtDNA). The sequenced region is the most variable on mtDNA molecule and is commonly used for population and evolutionary studies. Russians form one of the largest ethnic groups (more than 129 million). However, their genetic diversity had only been characterized with RFLP and biochemical markers, although there are already established mtDNA sequence databases for many ethnic groups of the world. We have obtained sequence data from 103 individuals living in three Russian regions: Kostroma, Kursk, and Rjazan. The sequenced fragment analyzed is 360 bp in length (positions from 16024 to 16383). Fifty nine nucleotide positions have been found polymorphic in Russians, among those were 57 transitions and two transversions. One individual is found having two insertions of two cytosines between positions 16184 and 16193. Among 64 different mitotypes identified in the study 52 were unique in these samples. The index of genetic diversity (Nei, 1987) for Russians is 0.96. This value is within the established range for European populations (0.93 to 0.98). Genetic distances calculated from our data show that Russians form a cluster with Germans, Bulgarians, Swedes, Estonians, and Volgo-Finns are more distant from Karelians and Finns, and much more differ from Turks and especially Mongolians.

Key words: mtDNA; Polymorphism; Russian

1. Introduction

Mitochondrial DNA (mtDNA) take a special place among high polymorphic genetic systems. The known nucleotide sequence, high level of variability and matrilineal kind of heredity allow to use mtDNA as an universal instrument in solving evolutionary and population genetic problems. Thus, mtDNA represents one of the most informative systems for inter- and intraspecific study of human genetic diversity.

The main non-coding region (control region) is a most variable part of mtDNA. The region contains origin of mtDNA replication and promoters for transcription [1]. Differences between individuals in the region are mainly presented by single nucleotide substitutions. More than half of these substitutions are located in the hypervariable segment I (HVSI). The comparative analysis of this segment was used in many population studies.

Nucleotide sequences of HVSI mtDNA have been obtained for many human populations in the world. The data on mitotype diversity of various ethnic and regional groups are widely used for studies of descendance and differentiation of ethnic groups, reconstruction of migrations and demographic processes.

Comparatively low level of pairwise differences and the specific geographic profile of mitotype distributions show that the European populations have a common ancestor [2]. However the ethnic history of the Eastern Slavs is not clear. It is supposed that Eastern Slavs could assimilate local tribes during outspreading into new areas.

Anthropological studies indicate the presence of the Finno-Ugric and Baltic components in ethnogenesis of the Eastern Slavs [3]. Thereby there is a great interest in revealing mtDNA diversity in Russians and comparing the data with those known for other populations from Europe.

The published data of mtDNA polymorphism in Russians are mainly based on analyses of restriction fragment length polymorphism (RFLP) [4–7]. However, resolution of RFLP is limited to the part of sequence presented in the recognition sites, and thus not all mitotypes can be revealed by this method. As for mtDNA sequences, there have not been published data except for a small sample (24 individuals) that was not sufficiently informative since it represented a mixture from different regions of Russia [8]. In this study we present data on nucleotide sequences of HVSI mtDNA region for 103 individuals from three regions of Russia and compare them to some close ethnic groups.

2. Materials and methods

The blood samples from 103 Russian individuals were collected on the territory of three regions in European part of Russia: Kostroma ($N=55$), Ryazan ($N=14$) and Kursk ($N=34$).

DNA samples were prepared as follows. 200 μl of whole blood were mixed with 300 μl of a buffer (100 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 mM NaCl, 1% SDS and 0.2 μg/ml proteinase K) and incubated at 37°C for 12–16 h. Then DNA was purified with phenol-chloroform extraction and concentrated by ethanol precipitation [9].

The following primers were used for amplification of HVSI: L15997: 5′-CACCATTAGCACCCAAAGCT-3′ and H16401: 5′-TGATTTCACGGAGGATGGTG-3′ [10], where 'L' and 'H' refer to 'light' and 'heavy' strands of mtDNA respectively and numbers indicate the 3′ base according to the Cambridge reference sequence of human mtDNA [1]. Amplification was carried out in 25 μl of reaction mixture, containing 25 mM Tris-HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 5 pmol of each primer, 100–200 ng of total DNA and 0.5 units of Taq polymerase. The amplification conditions (30 cycles) were the following: 94°C for 1 min, 50°C for 1 min, 72°C for 1 min, followed by final 5 min extension at 72°C.

PCR products were purified on 2% agarose gel followed by extraction with 'QIAEX II' kit (Qiagen).

*Corresponding author. Fax: +7 (95) 1351289.
E-mail: yankovsky@vigg.ru

PII: S0014-5793(99)00115-5

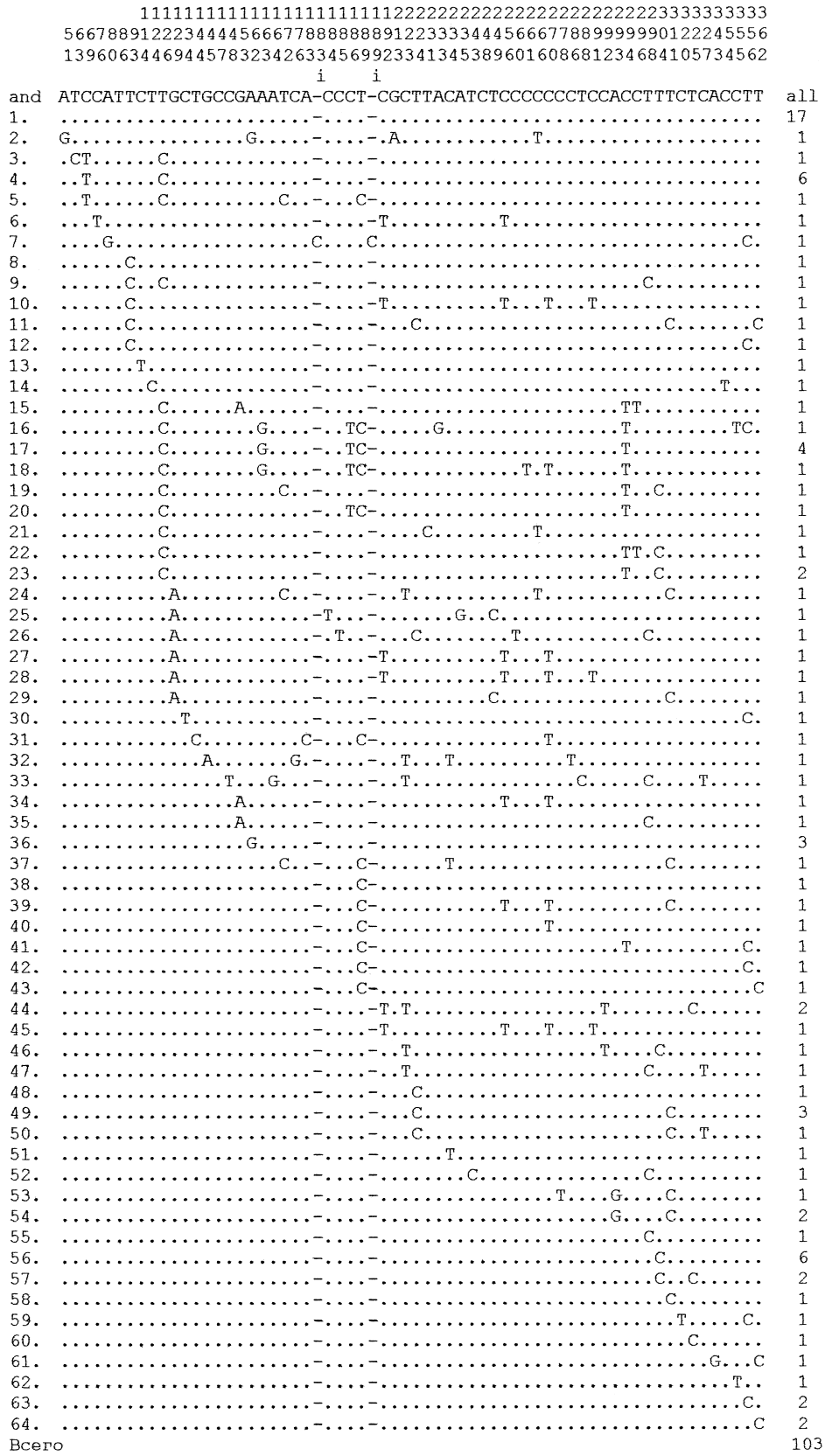

Fig. 1. Identified types of HVSI mtDNA control region. Only variable sites are shown. Numbers of positions shown according Cambridge reference sequence of human mtDNA without prefix '16' [1].

Table 1
Haplogroup frequencies (%) calculated for Russians compared to the published data for Europe in whole (Richards, 1996)

Haplogroup	Russians	Europe
1	54	62
2	20	20
2A	8	12
2B	12	8
3	4	5
4	5	7
5	7	7
Other	10	

Sequencing of the purified PCR products was performed on Applied Biosystems 373A DNA sequencer using 'ABI Prism' kit, containing FS polymerase, according to the manufacturer's protocol. L15997 oligonucleotide (see above) was used as the sequencing primer in most cases. In some cases (transition in position 16189) H16401 oligonucleotide (see above) was used as primer.

Published data for Bulgarians [11], Finns, Karelians, Volgo-Finns, Estonians, [12], North Germans, Swedes [13], Turks [11,14], and Mongolians [15] have been used in this study for comparison.

Mitotype diversity (H) can be estimated by $H=(1-\Sigma x^2)N/(N-1)$ [28], where x is the frequency of each mitotype and N the number of individuals, i.e. nucleotide sequences. The average pairwise difference, $\hat{\imath}$, has been calculated on the base of all $N(N-1)/2$ possible pairwise comparisons using software DNASP 2.2. The mean number of pairwise differences between all possible pair sequences (D_i, see below) may be used as a different measure of genetic diversity within a population [10,16].

Genetic distances between populations were estimated as $D=D_{ij}-(D_i+D_j)/2$, where D_i (D_j) is the within population diversity, the average value of pairwise differences between sequences drawn from population i (j), and D_{ij} is the between population diversity, the average of pairwise differences between sequences drawn from each population i and j. The dendrogram was constructed using UPGMA algorithm with SYSTAT program.

3. Results

Nucleotide sequences of HVSI (16024–16383) were obtained for 103 Russians individuals. Among 360 sites of the sequenced segment, 59 were found to be polymorphic. Of the variable sites, 57 nucleotide substitutions were transitions. Transversions were observed at two sites. The estimated transition:transversion ratio was 29:1, this agrees with other data obtained for humans [10]. Interestingly, one individual had insertion of two cytosines in region 16184–16193; the length polymorphism in this region has been already described earlier [17].

Sixty four different mitotypes have been identified among the 103 individuals. Fifty two of them (81%) were unique (Fig. 1). The latter value corresponds well to that for European

Table 2
Genetic distances between Russians (from our data) and some other populations

Ethnic group	Genetic distance with Russians
Swedes	−0.00006
Estonians	−0.00005
Volgo-Finns	−0.00003
Bulgarians	−0.00003
Germans	0.00003
Karelians	0.00005
Finns	0.00010
Turks	0.00122
Mongolians	0.00221

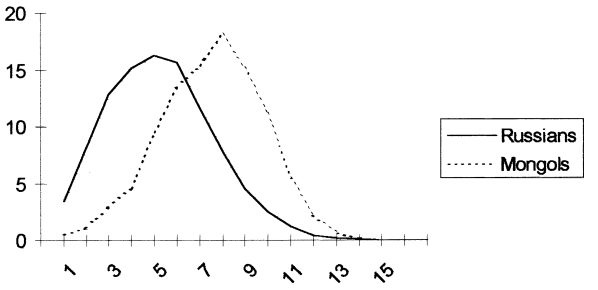

Fig. 2. Distribution of pairwise nucleotide differences for Russians and Mongols.

populations, 77% (Richards, 1996). The most frequent mitotype (16.5%) for the studied region of mtDNA was the Cambridge reference sequence [1]; in most European populations this mitotype has close frequencies [2].

In the previous article Richards et al. [2] offered classification of European mitotypes using the reduced median network. This classification is based on pooling mitotypes in five major lineage groups. In the examined group of Russians, the haplogroups defined by Richards et al. [2] occurred with frequencies close to those obtained for Europeans (Table 1).

Three mitotypes of Russian samples were previously shown to be presented in Mongolian populations. Mitotypes 33 and 47 belong to Mongoloid haplogroup C [19] and occurred in different Mongoloid populations [18,15]. Mitotype 26 was not found in Mongoloid populations; however, it probably originated from a Mongoloid mitotype because it contains substitutions similar to those for Mongolian mitotypes; at the same time it substantially differs from European mitotypes. The presence of these mitotypes in the examined Russian samples might indicate the (direct or indirect) influence of Mongoloid populations on the ethnogenesis of Russians.

The ethnospecific mitotypes were also found in Finno-Ugric populations [12]. These mitotypes have transitions in positions 16144, 16189, 16270 compared to the Cambridge reference sequence and have different frequencies in different Finno-Ugric populations. For example, in Finns this combination of transitions occurs with frequency 2%, in Karelians with frequency 6%. In other populations of the world this motif was not found. The analyzed group of Russians contains one 'Finno-Ugric' mitotype (mitotype 31). This confirms anthropological data about the contribution of Finno-Ugric populations in the ethnogenesis of Eastern Slavs [3].

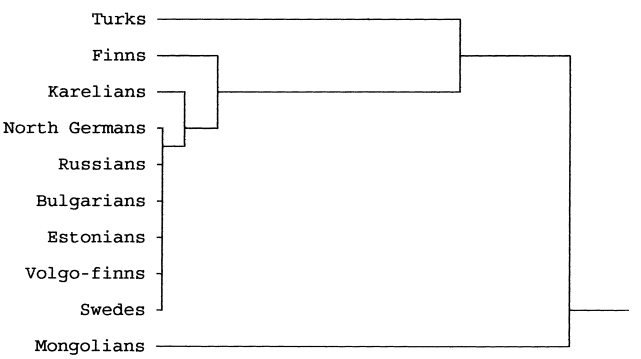

Fig. 3. UPGMA tree, built on a base of obtained genetic distances. The length of branches corresponds to genetic distance value.

Table 3
Restriction site frequencies of control region HVSI for Russians from different regions of Russia compared with Europeans in whole

Enzyme	Restriction site position and polymorphism type	Eastern Slavs of Magadan city [6] ($N=71$)	Russians of center and south of Russia [5] ($N=122$)	Russians of Siberia [7] ($N=126$)	Old Believers of Siberia [7] ($N=130$)	Europe [2,27] ($N=520$)	Our data ($N=103$)
*Eco*RV	(+)16274	14.1(10)	8.9(11)	6.4(8)	6.9(9)	1.9(10)	1.0(1)
*Hae*III	(+)16254	0.0(0)	0.8(1)	1.6(2)	0.0(0)	2.9(15)	0.0(0)
*Kpn*I	(−)16129	7.0(5)	14.8(18)	10.3(13)	12.2(16)	4.8(25)	5.7(6)
*Rsa*I	(−)16126	32.4(23)	–	7.9(10)	16.8(22)	20.2(105)	21.2(22)
	(−)16156	0.0(0)	1.6(2)	0.0(0)	0.0(0)	0.4(2)	0.0(0)
	(−)16208	0.0(0)	2.5(3)	1.6(2)	0.8(1)	0.4(2)	0.0(0)
	(−)16303	12.7(9)	10.6(13)	10.3(13)	6.9(9)	5.6(29)	12.5(13)
	(−)16311	12.7(9)	–	7.2(9)	14.3(19)	12.9(67)	12.5(13)
*Sau*3AI	(+)16145	1.4(1)	0.0(0)	–	–	–	0.0(0)
	(+)16215	0.0(0)	3.3(4)	–	–	–	0.0(0)

Mitotype diversity in the total sample of 103 individuals occurred to be 0.96, which is within the range for European populations (from 0.93 to 0.98 [2]). Another measure of population diversity, the mean value of nucleotide differences, calculated for all Russian individuals was 4.27 and is also within the range for Europeans [2].

The distribution of pairwise differences for different populations can be different and reflects some specific ancient demographic processes in these populations [20–22]. It is theoretically proven that a unimodal distribution of pairwise differences occurs in a population that had exponential growth periods in its demographic history. If a population had constant size over long time the distribution is expected to be multimodal [23]. This distribution in the analyzed Russian samples is unimodal (Fig. 2) and indicates the presence of population growth in the preceding history. It should be noted that the distribution found previously for Eastern Slavs was bimodal [24]. However, that sample was rather small (18 individuals) and such a bimodality could not be significant; additionally, it was collected in Magadan city among migrants from quite different regions of Russia [24] and thus was a mixture of possibly different distributions. The unimodal distribution is common for most European populations. This might be related with their active population growth in the upper paleolith [20]. Since the modal value of pairwise differences correlates with the duration of exponential population growth [20], it can be drawn from data that Mongolians started to grow much earlier than Russians (Fig. 2).

For estimating the genetic position of Russians among other ethnic groups of Europe and Asia genetic distances between Russians and some other populations have been calculated (Table 2) and interpreted by constructing a dendrogram (Fig. 3). The estimated genetic distances between Russians and most European populations were insignificant except for Karelians and Finns; the latter occurs to be genetically distinct from other Europeans [12].

A different way of data analysis is the use of frequencies of restriction sites. They have been calculated on the base of nucleotide sequences obtained and then compared with previously published data for Russians and Europeans [5–7]. The data (Table 3) show a tendency to heterogeneity among European populations for polymorphic restriction site (−)16126 *Rsa*I. Interestingly, there is a significant heterogeneity among Russian populations at this site ($\chi^2=12.1$ with d.f. = 7). Therefore there exist some differences between Russian regions in mtDNA pools.

4. Conclusion

According to the above mtDNA diversity analysis, Russians of the Eastern-European plain are close to the European ethnic groups; the same conclusion was obtained by using biochemical markers [25,26]. Additionally, our analysis shows some influence of Finno-Ugric ethnic group on the Russian mitochondrial pool. Our data also show that Russian populations are heterogeneous, this corresponds to anthropological data [2]. Further mtDNA analysis of samples from other regions of Russia is needed to understand the genetic structure of Russian populations.

Acknowledgements: The research was supported by a grant from Russian Foundation For Basic Research (# 98-04-48740) and a grant from Russian National Human Genome Program (# 39/98). We are thankful to Irina A. Lebedeva for help in experimental work, and to Prof T.I. Alekseeva for useful discussion of the paper.

References

[1] Anderson, S., Bankier, A.T. and Barrel, B.G. et al. (1981) Nature 290, 457–465.
[2] Richards, M., Corte-Real, H. and Forster, P. et al. (1996) Am. J. Hum. Genet. 59, 185–203.
[3] Alexeeva, T.I. (1973) Ethnogenesis of Eastern Slavs.
[4] Lemza, S.V. and Sokolova, O.V. (1992) Genetika 28, (5) 136–140.
[5] Petrischev, V.N. and Kutueva, A.B. (1993) Genetika 29, (8) 1382–1390.
[6] Derenko, M.V. and Maliarchuk, B.A. (1996) Genetika 32, (6) 815–821.
[7] Kazakovtseva, M.A., Voevoda, M.I., Babenko, V.N. and Osipova, L.P. (1998) Genetika 34, (4) 544–550.
[8] Maliarchuk, B.A., Derenko, M.V. and Solovenchuk, L.L. (1995) Genetika 31, (6) 846–851.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[10] Ward, R.H., Frazier, B.L., Dew, K. and Paabo, S. (1991) Proc. Natl. Acad. Sci. USA 88, 8720–8724.
[11] Calafell, F., Underhill, P. and Tolun, A. et al. (1996) Ann. Hum. Genet. 60, 35–49.
[12] Sajantila, A., Lahermo, P. and Anttinen, T. et al. (1995) Genome Res. 5, 42–52.
[13] Richards, M., Forster, P. and Tetzner, S. et al. (1995) Odens Suppl. 12, 141–163.
[14] Comas, D., Calafell, F. and Mateu, E. et al. (1996) Mol. Biol. Evol. 13, 1067–1077.
[15] Kolman, C.J., Sambuughin, N. and Bermingham, E. (1996) Genetics 142, 1321–1334.

[16] Bertranpetit, J., Sala, J. and Calafell, F. et al. (1995) Ann. Hum. Genet. 59, 63–82.
[17] Horai, S. and Hayasaka, K. (1990) Am. J. Hum. Genet. 46, 828–842.
[18] Derenko, M.V. and Shields, G.F. (1997) Mol. Biol. (Mosk.) 31, (5) 784–789.
[19] Torroni, A., Sukernik, R.I. and Schurr, T.G. et al. (1993) Am. J. Hum. Genet. 53, 591–608.
[20] Rogers, A.R. and Harpending, H. (1992) Mol. Biol. Evol. 9, 552–569.
[21] Slatkin, M. and Hudson, R.R. (1991) Genetics 129, 555–562.
[22] Harpending, H.C., Sherry, S.T., Rogers, A.R. and Stoneking, M. (1993) Curr. Anthropol. 34, 483–496.
[23] Mountain, J.L., Hebert, J.M. and Bhattagharyya, P.A. et al. (1995) Am. J. Hum. Genet. 56, 979–992.
[24] Maliarchuk, B.A. (1997) Genetika 33, (1) 101–105.
[25] Altukhov, Iu.P., Khil'chevskaia, R.I. and Shurkhal, A.V. (1981) Genetika XVII, (3) 548–555.
[26] Spitsyn, V.A., Afanas'eva, I.S., Agapova, R.K., Tsurikova, G.V., Shchekotikhina, Yu.A., Krause, D. and Kuchheuser, W. (1994) Genetika 30, (5) 702–708.
[27] Torroni, A., Lott, M.T. and Cabell, M.F. et al. (1994) Am. J. Hum. Genet. 55, 760–776.
[28] Nei, M. (1987) Molecular Evolutionary Genetics, Columbia University Press, New York.