

The Diversity of Y-Chromosome Lineages in Indigenous Population of South Siberia

M. V. Derenko^a, B. A. Malyarchuk^a, M. Wozniak^b, I. K. Dambuyeva^c, C. M. Dorzhu^d,
F. A. Luzina^e, H. K. Lee^f, D. Miscicka-Sliwka^{b†},
and Corresponding Member of the RAS I. A. Zakharov^g

Received July 27, 2005

DOI: 10.1134/S0012496606060123

Archeological and paleontological evidence suggests that South Siberia is an area where the most ancient contacts occurred between the members of Caucasoid and Mongoloid peoples. These contacts have substantially affected the racial type of most populations of Eurasia. Analysis of mitochondrial DNA (mtDNA) inherited without recombination in the maternal line has shown that Southern Siberian populations have developed on a heterogeneous genetic basis. This is a result of not only the diversity of Mongoloid components that were either indigenous to the gene pools of Siberian populations since the Paleolithic Age or introduced in different periods of time from Central Europe and East Asia, but also the presence of the Caucasoid component expressed in different degrees in most populations that have contributed into this heterogeneity [1, 2]. The nonrecombining portion of the Y chromosome inherited in the paternal line is another genetic system widely used for studying the population genetic history. The Y chromosome polymorphism has

been analyzed in a wide spectrum of Asian populations; nevertheless, the data on numerous aboriginal populations of Southern Siberia are scanty with respect to both the number of populations and the set of loci studied. Therefore, there is no comprehensive idea as to how the gene pools of individual ethnic groups have been formed, taking into account the contributions of both paternal and maternal lineages [3, 4].

The paternal lineages of the gene pools of South Siberian ethnic groups were characterized using the variation analysis of 17 Y chromosome diallelic loci (*DYS287*, *RPS4Y*, *SRY-8299*, *M89*, *M201*, *M52*, *M170*, *12f2*, *M9*, *M20*, *92R7*, *SRY-1532*, *DYS199*, *M173*, *M17*, *Tat* and *LLY22g*) that are polymorphic in different Eurasian populations [3–8]. The sample included 1358 subjects from ethnic groups of Siberia (Altaians-Kizhi, Teleuts, Shors, Tyvans, Tadjins, Tofalars, Sojots, Khakassians, Buryats, and Evenks), Central and East Asia (Mongolians and Koreans), and East Europe (Kalmyks and Russians) (Fig. 1). Our own data and those obtained for the similar set of markers by other authors were used for comparative analysis of Y chromosome variation in the East European populations of Russians (this study), Maris and Udmurts [8] (the total sample size was 503 subjects) and the Central and East Asian populations of Mongolians and Koreans (this study and [5, 6]), Kyrgyz, Dungans, Uighurs, Kazakhs, Uzbeks, Tajiks, Turkmen, northern Khani, Khueis, Manchu, and Orochens [5, 6] (in total, 959 subjects). Buryats have significant genetic similarity with the Mongolian-speaking Kalmyks, who originated from the western Oirat Mongols that populated the Baikal and upper Yenisei regions until the 12th century and then migrated to East Europe. Therefore, Buryats and Kalmyks were regarded as a “Siberian” group with respect to their origin.

On the basis of the variation of diallelic and microsatellite Y chromosome loci, the structure of the paternal lineages of Siberian ethnic groups was described in detail for the first time. We found a significant genetic differentiation between the ethnic groups of the Baikal and Altai–Sayan regions, which may be explained by

^a *Institute of Biological Problems of the North, Far East Division, Russian Academy of Sciences, ul. Portovaya 18, Magadan, 685000 Russia*

^b *Forensic Medicine Institute, Ludwik Rydygier Medicinal College, Nicolaus Copernicus University in Torun, ul. Sklodowskiej-Curie 9, 85-094 Bydgoszcz, Poland*

^c *Institute of General and Experimental Biology, Siberian Division, Russian Academy of Sciences, Ulan-Ude, 670047 Russia*

^d *Tyva State University, Kyzyl, 667003 Tyva Republic, Russia*

^e *Institute of Integrated Problems of Hygiene and Occupational Diseases, Siberian Division, Russian Academy of Medical Sciences, Novokuznetsk, 654051 Russia*

^f *College of Medicine, Seoul National University, Seoul, 151-742 Korea*

^g *Vavilov Institute of General Genetics, Russian Academy of Sciences, ul. Gubkina 3, Moscow, 113809 Russia*

[†] Deceased.

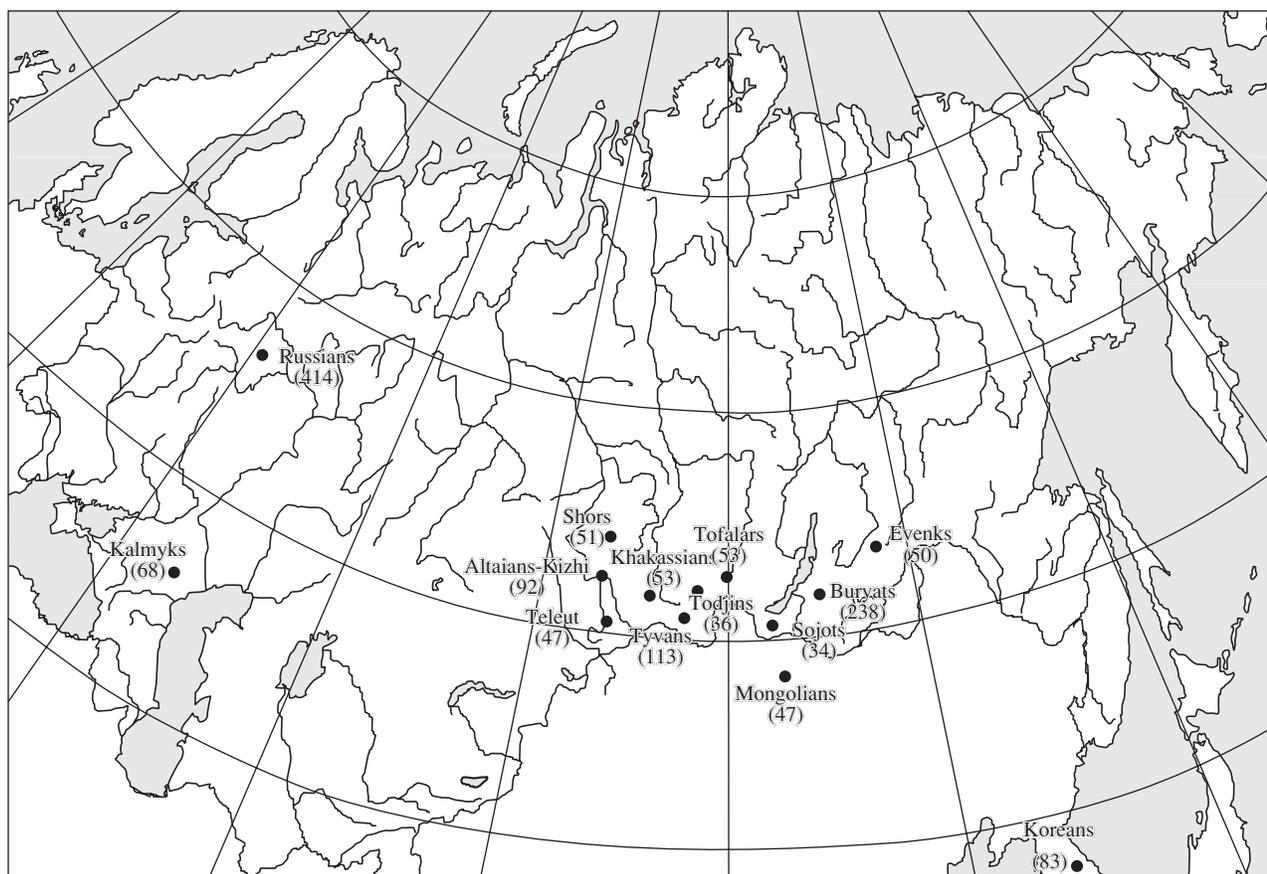


Fig. 1. Geographical location of the studied populations. The sample size is indicated in parentheses.

different contributions of the Central/East Asian and East European components into the gene pools of South Siberian ethnic groups. Phylogenetic relationships between the Y chromosome lineages within the R1a1 group were reconstructed for indigenous Siberian population and the Russian population of East Europe. The regional divergence of the Y-chromosome R1a1 lineages was shown to occur soon after the appearance of this haplogroup ($10\,310 \pm 3140$ years ago). Its evolutionary ages in the populations of both regions are similar: $11\,270 \pm 4070$ years in Siberia and $11\,380 \pm 3200$ years in Eastern Europe.

Our analysis showed that 15 out of 18 possible Y-chromosome groups were encountered in the studied populations (Table 1). In Siberian populations, the groups C, R1a1, N3, and P* were the most frequent; they make up 76% of the total sample. The AMOVA [10] of the Siberian population structure showed a high degree of interpopulation differentiation among indigenous populations of South Siberia ($F_{st} = 22.1\%$, $P = 0.000$), which is similar to the same parameter in Central/East Asia (24.7%, $P = 0.000$) and exceeds the value obtained for East Europe (13.8%, $P = 0.000$). The results of AMOVA also demonstrated that the genetic distances between South Siberian populations were sig-

nificantly correlated with geographical distances ($r = 0.77$; in the Mantel test, $P = 0.005$), which confirmed the decisive role of geographical factors in the differentiation characteristic of indigenous population of South Siberia.

The results of multidimensional scaling of the matrix of pairwise F_{st} distances (Fig. 2) also indicated interpopulation differentiation in the studied Asian region. For example, South Siberian populations formed three groups in the two-dimensional space of South Siberia: (1) Altai populations (Teleuts, Shors, and Altaians-Kizhi); (2) Sayan populations (Khakassians, Tyvans, Tadjiks, Tofalars) as well as Sojots; and (3) Baikal populations (Buryats, Kalmyks, and Evenks). Note that we observed a pronounced differentiation of the Altai–Sayan and Baikal regions with respect to the first dimension; with respect to the second dimension, Altai populations, which clustered together with the generalized East European group, distinctly segregated from Sayan and Baikal populations, which were genetically similar to the group of Central and East Asian populations. This differentiation may have been a consequence of different contributions of the components differing in their origin into the gene pools of the studied ethnic groups.

Table 1. The distribution of Y-chromosome groups (%) and genetic biodiversity in the studied populations

Population	N	P*	R1*	R1a1	N*	N3	BR*	DE*	C	F*	G	I	J	H	E*	K*	L	Genetic diversity \pm SE
Altaians-Kizhi	92	28.3	1.1	41.3	2.2	5.4	0	3.3	13.0	0	1.1	2.2	2.2	0	0	0	0	0.735 \pm 0.031
Teleuts	47	0	12.8	55.3	0	10.6	0	0	8.5	6.4	0	4.3	2.1	0	0	0	0	0.667 \pm 0.068
Khakassians	53	7.6	7.6	28.3	28.3	13.2	0	0	5.7	0	0	3.8	0	0	0	5.7	0	0.819 \pm 0.029
Shors	51	2.0	19.6	58.8	13.7	2.0	0	0	2.0	2.0	0	0	0	0	0	0	0	0.607 \pm 0.063
Todjins	36	22.2	2.8	30.6	2.8	11.1	0	0	8.3	2.8	0	2.8	0	0	2.8	13.9	0	0.838 \pm 0.036
Sojots	34	8.8	0	23.5	8.8	11.8	0	0	17.6	0	2.9	0	0	0	0	26.5	0	0.838 \pm 0.029
Buryats	238	1.7	0.8	2.1	1.3	18.9	0	0	63.9	1.7	0.4	0.4	0	0	0	8.8	0	0.550 \pm 0.033
Kalmyks	68	11.8	2.9	5.9	2.9	0	0	0	70.6	0	0	0	0	0	0	4.4	1.5	0.488 \pm 0.071
Evenks	50	0	6.0	14.0	18.0	16.0	0	0	40.0	4.0	0	2.0	0	0	0	0	0	0.772 \pm 0.040
Tofalars	32	3.1	12.5	12.5	34.4	25.0	0	0	6.3	0	0	3.1	0	0	0	3.1	0	0.807 \pm 0.043
Tyvans	113	35.4	0.9	17.7	14.2	9.7	0.9	0	7.1	3.5	0.9	0.9	0	0	0	8.9	0	0.807 \pm 0.023
Mongolians	47	4.3	4.3	2.1	6.4	2.1	0	0	57.4	0	0	2.1	0	0	0	21.3	0	0.629 \pm 0.067
Koreans	83	0	0	0	0	0	0	1.2	12.0	0	0	0	0	0	0	86.7	0	0.236 \pm 0.057
Russians	414	2.2	6.8	48.3	0.2	14.0	2.67	0	0.2	2.2	1.2	15.9	1.5	1.0	1.9	1.9	0	0.716 \pm 0.019

Note: N, sample size; SE, standard error.

The Y chromosome is characterized by a high degree of intercontinental differentiation [11], which makes it possible to use the data on the variation of Y chromosome markers for studying the mixing of populations and/or regional groups. We used our own and literature data to evaluate the contribution of the East European and Central/East Asian components into the gene pool of each of the ethnic groups. The mixing coefficient ($mY \pm SE$), which reflects the relative contribution of one of two potential components, was calculated for each population (Table 2) with the use of the Admix 2.0 software [12]. In the gene pools of Altaians-

Kizhi, Teleuts, and Shors, the maximum contribution of the East European Y-chromosome lineages was determined (70-98%). The East European component also significantly contributed (42-61%) to the gene pools of Tyvans, Tofalars, Todjins, and Khakassians. Conversely, the Buryat and Kalmyk gene pools contained exclusively the lineages of Central and East Asian origin. The same component was predominant in the gene pools of Evenks (72%) and Sojots (71%). Thus, clear-cut interregional distinctions are observed in the distribution of Y-chromosome lineages in South Siberia. The main East European contribution into the generalized gene pool of the modern South Siberian population is made by the populations of the Altai-Sayan region (67%), whereas in populations of the Baikal region, the Central and East Asian variants of Y chromosome predominate. Note that studies on the variation of mtDNA inherited in the maternal line [2] has also demonstrated that the frequency of Caucasoid lineages clinically decrease in a similar way, from the maximum in populations of the Alai-Sayan region as high as 34.5% in Altaians) to the minimum (less than 10%) in the population of the Baikal region.

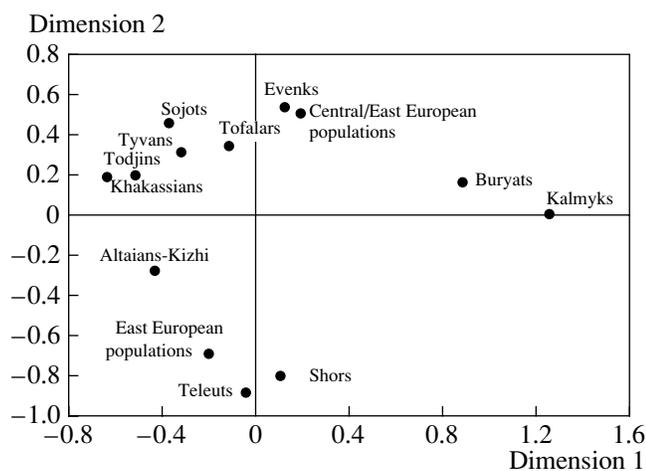


Fig. 2. Location of South Siberian, Central/East Asian, and East European populations in the two-dimensional space as determined by the multidimensional scaling of the matrix of F_{st} distances.

The R1a1 group is one of the genetic components shared by the male gene pools of European and Siberian populations. The high frequency of this group (up to 50%) was observed both in East Slavs (East Europe) and Altaians and Shors (South Siberia). The elevated frequency of the R1a1 group is known to be also characteristic of the eastern Iran population and some Indian populations. Analysis of the diversity of highly polymorphic microsatellite loci (STR loci) showed the highest diversity of R1a1 chromosome in southern populations of East Europe. Therefore, the most well-

grounded hypothesis thus far is that the R1a1 group has originated from the northern Black Sea area, and it has been spreading over different regions of Eurasia in the course of eastern European migration beginning from the Bronze Age or even earlier [7, 8]. Note, however, that, because of fragmentary data on the variation of STR loci of the Y chromosome in South Siberian populations, it is difficult to evaluate adequately the diversity of R1a1 lineages and the degree of divergence between the genetic pools of South Siberia and East Europe. In this study, the variation of 12 STR loci (*DYS19*, *DYS385a*, *DYS385b*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS437*, *DYS438*, and *DYS439*) was estimated, as well as the diversity of Y-chromosome lineages within the R1a1 group of South Siberian populations ($n = 134$) and in the Russian population of East Europe ($n = 121$). The diversity of R1a1-lineage haplotypes in Russians was found to exceed that in South Siberian populations ($h = 0.994$ and 0.959 , respectively); however, the degrees of diversity estimated from the differences in the number of repeats in all Y chromosomes and the ancient modal haplotype were almost the same ($ASD = 0.314$ and 0.313 , respectively). Phylogenetic analysis of the R1a1 lineages showed a pronounced differentiation of this component in the gene pools of South Siberia and East Europe: only 6% of STR haplotypes were shared by the compared samples (Fig. 3). This fact testifies to a sig-

Table 2. Relative contributions of the East European and Central/East Asian components ($mY \pm SE$) into the gene pools of South Siberian ethnic groups as estimated from the data on Y chromosome variation

Population	East European component	Central/East Asian component
Shors	0.983 ± 0.078	0.017 ± 0.078
Teleuts	0.972 ± 0.085	0.028 ± 0.085
Altaians-Kizhi	0.699 ± 0.070	0.301 ± 0.070
Khakassians	0.610 ± 0.103	0.390 ± 0.102
Todjins	0.507 ± 0.133	0.493 ± 0.133
Tofalars	0.502 ± 0.132	0.498 ± 0.132
Tyvens	0.419 ± 0.075	0.581 ± 0.075
Sojots	0.294 ± 0.134	0.706 ± 0.134
Evenks	0.281 ± 0.109	0.719 ± 0.109
Kalmyks	-0.207 ± 0.089	1.207 ± 0.089
Buryats	-0.078 ± 0.053	1.078 ± 0.053
Altai-Sayan region	0.670 ± 0.037	0.330 ± 0.037
Total of South Siberia	0.347 ± 0.034	0.653 ± 0.034

nificant interregional divergence of R1a1 chromosomes. Assuming that the mutation rate value was 0.00069 per generation [13], we calculated the evolu-

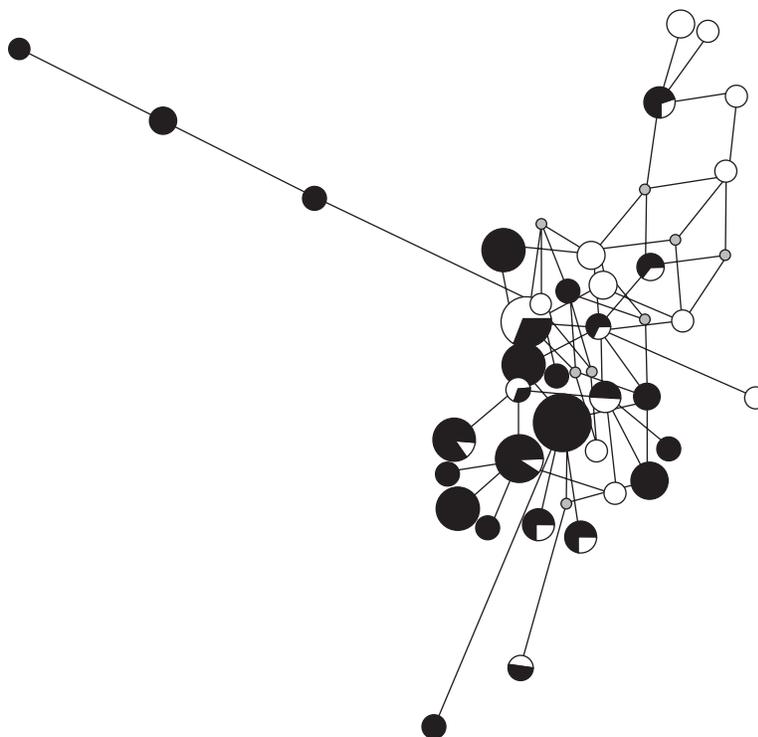


Fig. 3. The median network of STR haplotypes of the R1a1 group. The haplotypes detected in populations of South Siberia are marked in black; the haplotypes detected in the Russian population of East Europe are marked in white. The circle size is proportional to the haplotype frequency. The median vectors are grey. The network was constructed using the MJ (median joining) algorithm and the parameter “frequency > 1” [14].

tionary ages of the R1a1 component in Russian and South Siberian populations ($11\,380 \pm 3200$ and $11\,270 \pm 4070$ years, respectively) The time of divergence between these populations is $10\,310 \pm 3140$ years. Thus, our data suggest a very old age of the R1a1 lineages of the Y chromosome in the gene pool of indigenous South Siberian populations and a significant divergence of this component from that of the East European genetic pool. Further insight into the evolutionary history of the R1a1 haplogroup may be provided by new data on the variation of STR loci of the Y chromosome in the Iran Indian subcontinent populations.

In general, our results suggest a significant genetic differentiation between the ethnic groups of Baikal and Altai–Sayan regions, which is mainly accounted for by different contributions of the Central/East Asian and East European components into the gene pools of South Siberian ethnic groups.

ACKNOWLEDGMENTS

We are grateful to L.A. Zhyvotovskiy for the help in statistical analysis of the data on the variation of microsatellite loci of the Y chromosome.

This study was supported by the Russian Foundation for Basic Research (project no. 04-04-48746) and the Program of Fundamental Research of the Presidium of the Russian Academy of Sciences “Gene Pool Dynamics and Biodiversity.”

REFERENCES

1. Derenko, M.V., Malyarchuk, B.A., Dambueva, I.K., and Zakharov, I.A., *Dokl. Biol. Sci.*, 2003, vol. 393, no. 5, pp. 557–561 [*Dokl. Akad. Nauk*, 2003, vol. 393, no. 5, pp. 710–714].
2. Derenko, M.V., Grzybowski, T., Malyarchuk, B.A., et al., *Ann. Hum. Genet.*, 2003, vol. 67, pp. 391–411.
3. Karafet, T.M., Osipova, L.P., Gubina, M.A., et al., *Hum. Biol.*, 2002, vol. 74, pp. 761–789.
4. Stepanov, V.A., *Etnogenomika naseleniya Sibiri i Srednei Azii* (The Ethnogenomics of Siberian and Central Asian Populations), Tomsk: Pechatnaya Manufaktura, 2002.
5. Zerjal, T., Wells, R.S., Yuldasheva, N., et al., *Am. J. Hum. Genet.*, 2002, vol. 71, pp. 466–482.
6. Karafet, T.M., Xu, R., Du, R., et al., *Am. J. Hum. Genet.*, 2001, vol. 69, pp. 615–628.
7. Rosser, Z.H., Zerjal, T., Hurler, M.E., et al., *Am. J. Hum. Genet.*, 2000, vol. 67, pp. 1526–1543.
8. Semino, O., Passarino, G., Oefner, P.J., et al., *Science*, 2000, vol. 290, pp. 1155–1159.
9. Galushkin, S.K., Spitsyn, V.A., and Crawford, M.H., *Hum. Biol.*, 2001, vol. 73, pp. 823–834.
10. Schneider, S., Roessli, D., and Excoffier, L., *Arlequin. Vers. 2.000: A Software for Population Genetics Data Analysis*, Geneva: Univ. of Geneva, 2000.
11. Seielstadt, M.T., Minch, E., and Cavalli-Sforza, L.L., *Nat. Genet.*, 1998, vol. 20, pp. 278–280.
12. Dupanloup, I. and Bertorelle, G., *Mol. Biol. Evol.*, 2001, vol. 18, pp. 672–675.
13. Zhivotovskiy, L.A., Underhill, P.A., Cinnioglu, C., et al., *Am. J. Hum. Genet.*, 2004, vol. 74, pp. 50–61.
14. Bandelt, H.J., Forster, P., and Rohl, A., *Mol. Biol. Evol.*, 1999, vol. 16, pp. 37–48.