
**GENOMICS.
TRANSCRIPTOMICS**

UDC 575.174:599.9

Genetic Diversity and Relationships of Populations of Northern Eurasia by Polymorphic Alu Insertions

I. Yu. Khitrinskaya^a, V. N. Kharkov^a, M. I. Voevoda^b, and V. A. Stepanov^a

^a *Research Institute of Medical Genetics, Russian Academy of Medical Sciences, Tomsk, 634050 Russia;*
e-mail: vadim.stepanov@medgenetics.ru

^b *Research Institute of Internal Medicine, Russian Academy of Sciences, Novosibirsk, 630089 Russia*

Received June 20, 2013; in final form, July 16, 2013

Abstract—Autosomal gene pools of 27 populations representing 12 ethnic groups of Siberia, Central Asia, and the Far East have been characterized for the first time using a set of eight polymorphic Alu insertions. The results of our analysis indicate a significant level of genetic diversity in populations of northern Eurasian and the considerable differentiation of their gene pool. It was shown that the frequency of the Alu (–) allele at the CD4 locus was inversely related to the magnitude of the Mongoloid component of the gene pool: the lowest and highest frequencies of the CD4 Alu deletion were recorded in Eskimos (0.012) and in Russians and Ukrainians (0.35), respectively. A gene flow analysis showed that Caucasoid populations (Russians, Tajiks, and Uzbeks), as well as Turkic ethnic groups of southern Siberia (Altaians and Tuvans), Khanty, and Mansi populations, in contrast to ethnic groups of eastern Siberia and the Far East, have been recipients of a considerable gene flow. A correlation analysis showed that genetic distances determined using polymorphic Alu insertions were correlated with the anthropological characteristics of the populations studied.

DOI: 10.1134/S0026893314010051

Keywords: Alu-insertions, genetic diversity, populations of northern Eurasia

INTRODUCTION

Ethnogenetic studies aimed at characterization of gene pool structure of human populations are conducted at multiple research centers throughout the world. Various systems of genetic markers are employed to analyze genetic diversity and differentiation, as well as to reconstruct ancient migrations and the population and demographic history of particular regions or ethnic groups [1–12]. Populations of Siberia, Central Asia, and the Far East, which are highly diverse in their origins and ethnicity, are being extensively investigated using different systems of genetic markers, e.g., SNPs and microsatellites of the nonrecombining region of chromosome Y, mtDNA, or chromosome Ö haplotypes. However, it is highly desirable that the array of genetic markers in use be further extended, since Siberia and Central Asia have always been regions of contacts among a great variety of ethnocultural and anthropological entities from adjacent territories.

Although population genetics studies have lately been making increasingly extensive use of new genotyping approaches, such as genome-wide SNP analysis, new-generation sequencing, or complete mtDNA and genomic sequencing, classic DNA markers, such as polymorphic Alu repeats, are still in vogue. The high information content of these DNA markers, as well as the accuracy and simplicity of genotyping and the availability of large databases for comparative

analysis, makes them an appropriate tool for studying gene-pool structure at a level sufficient for most problems of population genetics.

Alu insertion polymorphisms are convenient markers for this type of studies. Several research centers in Europe and the United States investigate genetic diversity of human populations based on Alu repeat markers. The collected data made it possible to characterize human genetic diversity and to verify major pathways of human expansion over the Earth [13], as well as to specify genetic relationships among different Native American groups [14].

Alu insertions have been in use to study gene pools of northern Eurasian populations for more than 15 years. In this period, researchers characterized the general structure of gene pools of different ethnic groups, including eastern Slavic populations and indigenous peoples of the Volga–Urals region, the North Caucasus, and Siberia, analyzed phylogenetic relationships among these populations and determined the allele frequencies for different Alu markers. It was also possible to reconstruct historical genetic relationships among some of the indigenous peoples of Russia [2, 15–28].

However, the resulting picture still remains largely fragmented; it does not fully reflect the general organization patterns that characterize the gene pool of northern Eurasia. In addition, to a certain extent, many ethnic groups are anthropologically and geneti-



Fig. 1. Geographic location of ethnic groups studied. Here and in other figures, group notation is given according to Table 1.

cally heterogeneous, and population samples used in previous studies often underrepresented the total gene pool variation and only partially characterized its actual diversity. Therefore, although the body of accumulated data is considerable, we believe that it is important to investigate the regional and intragroup structure of genetic diversity of different populations based on at least a few geographically separate samples for each region.

EXPERIMENTAL

Populations. The study involved 34 population samples that represent 19 ethnic groups residing in Siberia, Central Asia, the Far East, and Eastern Europe (Fig. 1, Table 1), with a total of 2383 individuals. These ethnic groups represent two major Eurasian racial types (Caucasoid and Mongoloid) and speak languages from six different families, i.e., Altaic, Indo-European, Sino-Tibetan, Uralic, Chukotko-Kamchatkan, Eskimo-Aleutan, and Paleosiberian.

DNA polymorphism was studied using eight Alu insertion loci, including ACE, FXIIIIB, APOA1, PLAT, PV92, A25, CD4, and D1. DNA was isolated using conventional techniques [29]. Genotyping was performed by PCR with subsequent electrophoresis in

a 2% agarose gel. Statistical analysis was performed using conventional approaches of population genetics studies [30–33].

RESULTS AND DISCUSSION

Allele Frequencies and Genetic Diversity within Populations

All eight loci studied were polymorphic in all populations (Table 2); four of them showed a high level of diversity; i.e., for ACE, PLAT, PV92, and D1, the H_e value in the total sample was close to 0.5. In general, the spectrum of allele frequencies in indigenous populations of Siberia was typical for other Asian populations; the frequencies of Alu insertions at PV92 and F13B were relatively high [13, 14, 34]. Populations of Siberia and Central Asia can be divided in two groups depending on F13B allele frequencies. The first group comprises Caucasoid populations (Russians, Ukrainians, Tajiks, Uzbeks) and those of the intermediate Uralic type (Khanty and Mansi), while the other one includes all Mongoloid ethnic groups.

In populations of Central Asia, allele frequency distributions were heterogeneous. The Kyrgyz and Dungans were characterized with typical Mongoloid allele distribution patterns, Tajiks were closer to Cau-

Table 1. Linguistic and anthropological characteristics of the populations studied

Ethnic group	Population (settlement)	Notation	N*	Location	Linguistic classification (family/group)	Racial and anthropological type
1	2	3	4	5	6	7
Altaians, Northern	Gorno-Altaiisk	Alt-GA	31	Republic of Gornyi Altai	Altaic/Turkic	Mongoloid (South Siberian)
	Kurmach-Baigol	Alt-KB	29	Republic of Gornyi Altai		
Alatians, Southern	Kulada	Alt-KU	93	"		Altaic/Turkic
	Kosh-Agach	Alt-KA	43	"		
	Beshpel'tir	Alt-BE	114	"		
Tuvans	Teeli	Tuv-TE	129	Republic of Tuva	Altaic/Turkic	Mongoloid (Central Asian)
	Kungurtug	Tuv-KT	165	Republic of Tuva		
	Toora-Khem	Tuv-TH	113	"		
Buryats	Ulan-Ude	Bur-UU	60	Republic of Buryatia	Altaic/Mongolian	Mongoloid (Central Asian)
	Khuramsha	Bur-HU	60	Republic of Buryatia		
	Kurumkan	Bur-KU	88	"		
	Aginskoe	Bur-AG	78	Chita oblast		
Evenks	Chara	Evk-CH	42	Chita oblast	Altaic/Manchu-Tungusic	Mongoloid (Baikalian)
	Tungokochen	Evk-TN	11	"		
	Olekma (Tupik and Maklakan)	Evk-OL	42	"		
Yakuts	Cheriktei	Yak-CH	81	Sakha	Altaic/Turkic	Mongoloid (Central Asian)
	Dyupsya	Yak-DY	64	Sakha		
	Byadi	Yak-BY	56	"		
Kets	Kellog	Kek	40	Krasnoyarsk krai; Turukhanskii district	Paleosiberian/Ket	Mongoloid (Baikalian)
Kyrgyz	North (Kegety, Taldy-Su)	Kir-N	50	Kyrgyzstan	Altaic/Turkic	Mongoloid (South Siberian)
	South (Osh)	Kir-S	54	Kyrgyzstan		
Uzbeks	Osh	Uzb	46	Kyrgyzstan	Altaic/Turkic	Caucasian (Pamir-Iranian)
Tajiks	Dushanbe	Taj	48	Tajikistan	Indo-European (Iranian)	Caucasian (Pamir-Iranian)
Dungans	Tokmak, Aleksandrovka	Dun	48	Kyrgyzstan	Sino-Tibetan/Chinese	Mongoloid (East Asian)
Russians	Tomsk	Rus-T	112	Tomsk	Indo-European/Slavic	Caucasian (East European)
	Kargala	Rus-K	103	Tomsk oblast		
Kazakh	Kosh-Agach	Kaz	80	Republic of Gornyi Altai	Altaic/Turkic	Mongoloid (Central Asian and South Siberian)
Ukrainians	Kharkov, Poltava	Ukr	97	Ukraine	Indo-European/Slavic	Caucasian (East European)

Table 1. (Contd.)

Ethnic group	Population (settlement)	Notation	N^*	Location	Linguistic classification (family/group)	Racial and anthropological type
1	2	3	4	5	6	7
Chukchi	Lorino, Yanarykot, Novoe Chaplino, Sireniki	Chuk	60	Chukotka Autonomous Okrug	Chukotko-Kamchatkan	Mongoloid (Arctic)
Khanty	Tyumen oblast	Khant	79	Tyumen oblast	Uralic/Finno-Ugric	Uralic
Mansi	Tyumen oblast	Mans	40	Tyumen oblast	Uralic/Finno-Ugric	Uralic
Koryaks		Kor	71	Koryak Autonomous Okrug, Kamchatka	Chukotko-Kamchatkan	Mongoloid (Arctic)
Eskimo	Igloolik	Eskim	123	Canada	Eskimo-Aleut	Mongoloid (Arctic)
Nivkh	Southern Sakhalin	Nivh	33	Sakhalin oblast	Paleosiberian/nivkh	Mongoloid (Sakhalin-Amur)

* N is number of individuals studied.

casoid populations, whereas Turkic-speaking Uzbeks had intermediate allele frequencies that tend toward Caucasoid-type distributions. An interesting phenomenon was observed for the CD4 locus, i.e., the frequency of the deletion allele was clearly decreasing with an increase in the Mongoloid component of the gene pool. Among the populations of Siberia and Central Asia included in our study, the lowest frequency of the Alu deletion at CD4 was found in the Eskimo population (0.012), which has the strongest Mongoloid component, while in Russians and Ukrainians, the frequency of this deletion was the highest (0.35).

The highest level of average expected heterozygosity by the eight Alu loci was observed in Northern Altaians from Kurmach-Baigol ($H_e = 0.4195$) and Gorno-Altaiisk ($H_e = 0.4091$). In Tuvans, the expected heterozygosity was also high, which, as shown previously in [35], is related to heterogeneity of this ethnic group. Among the three Tuvan populations studied, the lowest diversity was observed in the highly isolated population of Toora-Khem [36, 37], which is in agreement with the data on the spread and accumulation of family names (tribal groups) [35]. Our data suggesting the low genetic diversity of the Toora-Khem population agree with the results of an RFLP analysis of mtDNA D-loop, but disagree with the data obtained using classic markers, which detected the highest level of genetic diversity in the Todzha population [35].

Genetic Differentiation among Populations

The level of diversity among populations was evaluated using the genetic differentiation coefficient G_{ST} , which reflects the portion of interpopulation differences in the total genetic diversity within a group of populations (Table 3). The highest contribution to interpopulation diversity was associated with differ-

ences in Alu insertion frequencies at PV92 ($G_{ST} = 12.6\%$), APOA1 ($G_{ST} = 12\%$), F13B ($G_{ST} = 11.7\%$), and CD4 ($G_{ST} = 10.1\%$). There was little differentiation among the populations studied by the PLAT and A25 markers ($G_{ST} = 2.1$ and 2.2% , respectively).

On the whole, the gene pool of populations that represent different geographic regions and linguistic groups of northern Eurasia was highly differentiated: the mean G_{ST} value was 7.5% . This level of interpopulation genetic differences reflects well the complexity of the ethno-population system in question. For comparison, the genetic differentiation among 34 populations representing all major geographic regions of the world determined using polymorphism of Alu insertions was 12.8% [13]. Importantly, very similar levels of genetic differentiation among human populations are observed using large random sets of autosomal DNA markers, including genome-wide panels of hundreds of thousands SNPs. For instance, when 650000 SNPs were genotyped in 51 populations from the panel of the Human Genome Diversity Project, it was found that interpopulation differences account for 11% of the worldwide genetic diversity [38]. Our recent evaluation of genetic differentiation among 36 human populations (32 Eurasian populations and 4 populations from the HapMap project) using 200000 SNPs produced a value of 13.4% [39].

Next, we analyzed the differentiation within individual ethnic groups. Among the populations of Siberia and Central Asia included in our study, the highest G_{ST} value was observed in Yakuts (2.1%). In Buryats and Altaians, the portion of interpopulation differences was slightly lower (1.98 and 1.7% , respectively), while in other ethnic groups, it did not exceed 1.5% . These estimates show that gene pools of indigenous ethnic groups (Tuvans, Evenks, Kyrgyz) are internally homogeneous, which can probably be traced back to

Table 2. Allele frequencies and heterozygosity levels by eight polymorphic Alu insertion loci

Population	N	ACE		PLAT		PV92		APOA1		F13B		A25		CD4		D1	
		f	He														
Kungurtug (Tuv-KT)	165	.512	.499	.397	.478	.685	.431	.684	.432	.795	.325	.119	.209	.071	.132	.675	.438
Teeli (Tuv-TE)	114	.385	.473	.406	.482	.614	.473	.774	.349	.715	.406	.089	.163	.088	.160	.477	.499
Toora-Khem (Tuv-TH)	129	.546	.495	.415	.485	.700	.420	.693	.424	.881	.209	.101	.183	.063	.118	.709	.411
Tuvans, total	407	.487	.499	.404	.481	.665	.445	.718	.404	.792	.328	.103	.185	.074	.137	.628	.467
Ulan-Ude (Bur-UU)	60	.691	.426	.475	.498	.683	.432	.912	.160	.847	.258	.093	.169	.025	.050	.342	.450
Kurumkan (Bur-KU)	88	.545	.493	.462	.495	.715	.406	.867	.229	.789	.332	.073	.136	.074	.138	.675	.438
Khuramsha (Bur-HIR)	59	.593	.482	.533	.497	.658	.449	.898	.182	.825	.288	.075	.138	.050	.095	.560	.492
Aginskoe (Bur-AG)	77	.653	.452	.435	.491	.750	.375	.916	.152	.883	.206	.109	.194	.044	.085	.628	.466
Buryats, total	284	.618	.473	.472	.498	.706	.414	.896	.185	.834	.276	.087	.160	.051	.097	.567	.490
Chara (Evk-CH)	41	.646	.457	.428	.489	.595	.481	.892	.191	.780	.427	.075	.138	.050	.095	.576	.488
Olekma (Evk-OL)	42	.702	.418	.500	.500	.536	.497	.952	.090	.841	.266	.134	.232	.024	.047	.735	.389
Tungokochoen (Evk-TN)	11	.681	.433	.454	.495	.681	.433	.909	.165	.818	.297	.045	.086	.090	.165	.681	.433
Evenks, total	94	.675	.438	.463	.497	.579	.487	.921	.145	.811	.305	.097	.176	.043	.083	.654	.452
Kulada (Alt-KU)	114	.657	.450	.447	.494	.610	.485	.886	.202	.769	.354	.070	.130	.048	.091	.675	.438
Beshpel'tir (Alt-BE)	93	.629	.166	.505	.499	.531	.498	.849	.255	.693	.425	.053	.100	.118	.208	.738	.386
Kosh-Agach (Alt-KA)	43	.581	.486	.558	.493	.662	.447	.918	.149	.812	.304	.081	.149	.104	.187	.445	.494
Gorno-Altaiisk (Alt-GA)	31	.564	.491	.435	.491	.580	.487	.725	.398	.741	.382	.080	.148	.274	.398	.611	.475
Kurmach-Baigol (Alt-KB)	29	.637	.462	.706	.414	.448	.494	.706	.414	.741	.385	.137	.237	.396	.478	.620	.470
Altaians, total	310	.624	.467	.503	.500	.575	.488	.846	.259	.746	.378	.074	.137	.132	.229	.641	.459
Kets (Kek)	40	.700	.420	.538	.497	.487	.499	.833	.277	.671	.441	.050	.095	.050	.095	.513	.499
Ukrainians (Ukr)	97	.453	.495	.526	.498	.234	.358	.911	.472	.382	.472	.123	.216	.350	.455	.456	.496
Cheriktei (Yak-CH)	81	.722	.401	.556	.493	.696	.423	.895	.187	.808	.309	.043	.082	.024	.048	.637	.462
Dyupsya (Yak-DY)	64	.578	.487	.523	.498	.698	.421	.951	.092	.835	.274	.070	.130	.033	.064	.550	.494
Byadi (Yak-BY)	56	.625	.468	.571	.489	.787	.335	.821	.293	.812	.304	.026	.052	.026	.052	.330	.442

Table 2. (Contd.)

Population	N	ACE		PLAT		PV92		APOAI		F13B		A25		CD4		DI	
		f	He	f	He	f	He	f	He	f	He	f	He	f	He	f	He
Yakus, total	201	.649	.455	.550	.495	.721	.401	.892	.192	.818	.292	.047	.090	.027	.054	.508	.499
Kyrgyz, South (Kir-S)	50	.610	.475	.470	.498	.500	.500	.857	.244	.829	.282	.080	.147	.091	.166	.771	.352
Kyrgyz, North (Kir-N)	54	.648	.456	.470	.498	.576	.488	.945	.103	.788	.333	.094	.170	.077	.143	.578	.487
Kyrgyz, total	104	.628	.466	.609	.476	.491	.499	.886	.200	.808	.310	.087	.159	.125	.218	.747	.377
Tomsk (Rus-T)	107	.429	.490	.575	.488	.176	.290	.894	.188	.433	.491	.111	.198	.350	.455	.495	.500
Kargala (Rus-K)	102	.524	.498	.543	.496	.210	.332	.953	.088	.391	.476	.087	.159	.356	.458	.450	.495
Russians, total	209	.471	.498	.560	.492	.193	.311	.922	.143	.413	.484	.100	.180	.352	.456	.475	.498
Uzbeks (Uzb)	46	.597	.480	.510	.499	.466	.497	.913	.158	.554	.494	.085	.155	.130	.226	.465	.497
Dungans (Dun)	44	.602	.479	.602	.479	.772	.351	.886	.201	.681	.433	.045	.086	.045	.086	.458	.496
Kazakh (Kaz)	79	.601	.479	.533	.497	.550	.495	.924	.140	.797	.323	.072	.134	.129	.226	.600	.480
Tajiks (Taj)	41	.329	.441	.585	.485	.537	.497	.902	.176	.622	.470	.128	.223	.146	.249	.666	.444
Chukchis (Chuk)	57	.640	.460	.406	.482	.722	.401	.982	.034	.925	.137	.158	.266	.035	.067	.627	.467
Eskimo (Esk)	113	.672	.440	.421	.487	.774	.349	.922	.143	.973	.051	.081	.149	.012	.025	.639	.461
Nivkh (Niv)	33	.697	.422	.393	.477	.900	.180	.984	.029	.893	.189	.075	.140	.075	.140	.700	.042
Khanty (Khnt)	79	.708	.412	.649	.455	.297	.417	.867	.230	.660	.448	.120	.211	.170	.283	.675	.438
Mansi (Mans)	38	.750	.375	.550	.495	.300	.420	.912	.159	.555	.493	.112	.199	.275	.398	.472	.498
Koryaks (Kor)	71	.816	.299	.528	.498	.507	.499	.993	.014	.928	.132	.042	.080	.042	.082	.470	.498

N is number of individuals; f is frequency of Alu deletions (- allele) at CD4 and Alu insertions (+ allele) at all other loci. He is heterozygosity of each population; initial zero is not shown.

Table 3. Average expected heterozygosity (H_S), interpopulation diversity (D_{ST}), and genetic differentiation coefficient (G_{ST}) for each polymorphic locus studied

Locus	H_T	H_S	D_{ST}	G_{ST} , %
ACE	0.4804	0.4536	0.0268	5.57
PLAT	0.4999	0.4893	0.0106	2.1
PV92	0.4910	0.4283	0.0629	12.6
APOA1	0.2246	0.1977	0.0269	12.0
F13B	0.3780	0.3336	0.0444	11.7
A25	0.1600	0.1565	0.0035	2.2
CD4	0.2021	0.1817	0.0204	10.1
D1	0.4842	0.4619	0.0223	4.6
Σ	0.3651	0.3378	0.0273	7.5

the ancestry of particular ethnic groups or to the history of their migrations.

The results of a hierarchical analysis of genetic differentiation among territorial, linguistic, and anthropological groups based on Alu insertion polymorphisms are shown in Table 4.

For the population of Siberia, which was represented by 21 populations included in this study, the genetic differentiation coefficient was 4.4%, which indicates a rather high level of genetic diversity within

the region. Studies employing immunological and biochemical markers also demonstrated a high level of diversity among Siberian populations [40]. The highest F_{ST} = 3.6% levels in Eurasia were observed in central higher than the average F_{ST} = 3.0% value for northern Eurasia [40].

Similar G_{ST} values were observed in the populations of Central Asia and eastern Siberia (2.7 and 2.2%, respectively). For thousands of years, Central Asia has been a scene of important genetic and demographic processes, which explains a considerable portion of interpopulation differences. Migration processes in eastern Siberia have been, in contrast, much less active. Anthropological relationships among Yakuts, Buryats, and Evenks are closer than among populations of Central Asia, but the lack of close contact among these ethnic groups due to territorial separation also resulted in accumulation of significant interpopulation differences.

In ethnic groups of southern Siberia (Altai and Sayan Mountains), the portion of interpopulation differences was low (less than 1%). Tuvans and Northern and Southern Altaians speak languages of the same Turkic group of the Altaic family; the gene pool of this region can probably be considered homogeneous. The gene pool of western Siberia, represented by two populations of closely related ethnic groups, Khanty and Mansi, was even more uniform (G_{ST} = 0.37%) [41].

Table 4. Genetic differentiation within territorial, linguistic, and anthropological groups of populations of northern Eurasia

Group	Number of populations	H_T	H_S	D_{ST}	G_{ST} , %
<i>Territorial groups</i>					
Siberia	21	0.3568	0.3413	0.0155	4.34
Southern	8	0.3702	0.3675	0.0027	0.72
Western	2	0.3725	0.3711	0.0014	0.37
Eastern	10	0.3205	0.3133	0.0072	2.25
Far East	4	0.2764	0.2665	0.0099	3.58
Central Asia	6	0.3584	0.3487	0.0097	2.70
Eastern Europe	3	0.3869	0.3857	0.0012	0.31
<i>Linguistic groups</i>					
Altaic family	22	0.3534	0.3408	0.0126	3.56
Turkic group	15	0.3613	0.3508	0.0105	2.91
Indo-European family	4	0.3921	0.3825	0.0096	2.44
Uralic family	2	0.3725	0.3711	0.0014	0.37
<i>Anthropological groups</i>					
Caucasoid racial type	5	0.3936	0.3815	0.0121	3.07
Mongoloid racial type	27	0.3402	0.3272	0.0013	3.82
Uralic racial type	2	0.3725	0.3711	0.0014	0.37

H_T is total genetic variation; H_S is variation among individuals within populations; G_{ST} is coefficient of genetic differentiation.

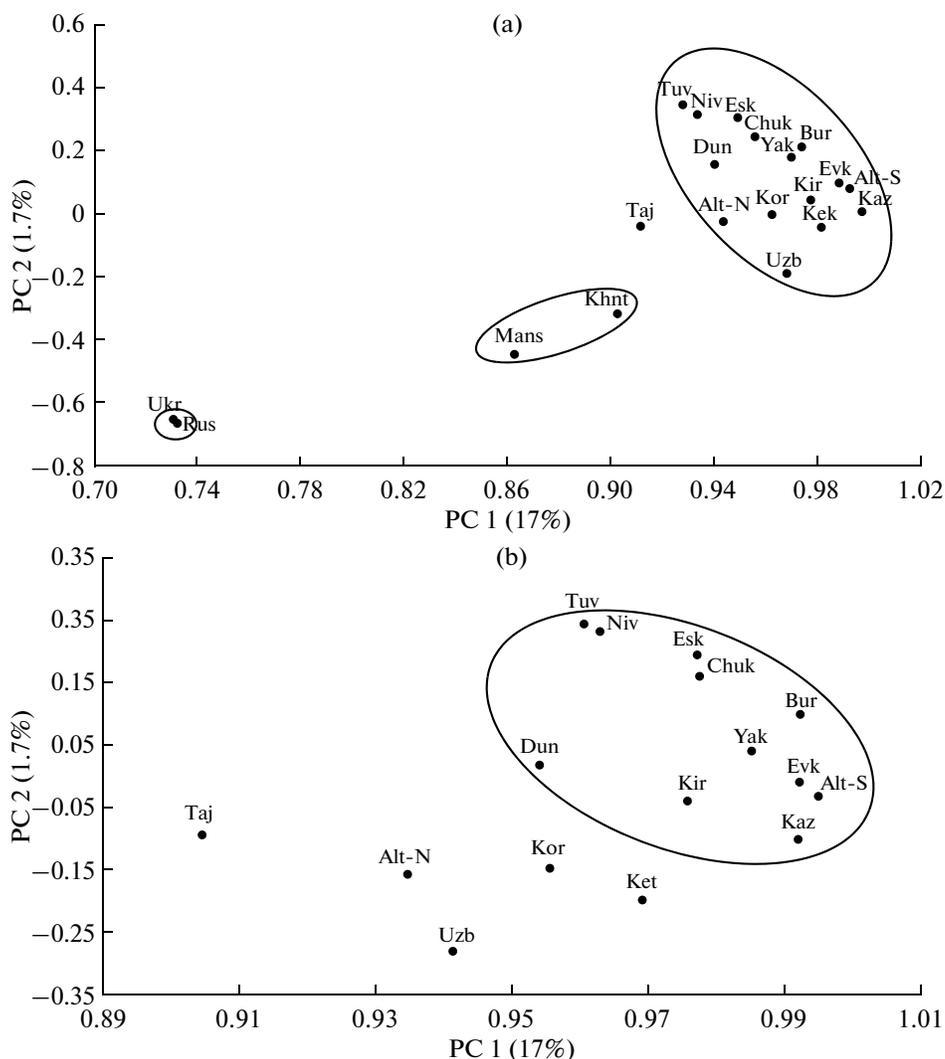


Fig. 2. Ethnic groups of Northern Eurasia in principal component space of allele frequencies. (a) 20 ethnic groups; (b) 16 ethnic groups.

For the Altaic language family, the genetic differentiation coefficient was 3.6%. Differentiation within the Turkic group of this family ($G_{ST} = 2.9\%$) accounted for 97% of the total differentiation within the family, the percentage that corresponds to the portion of the Turkic-speaking group in the total number of Altaic population.

The evaluation of interpopulation differences within anthropological groups showed that two large racial groups, Caucasoid and Mongoloid, had similar levels of genetic differentiation ($G_{ST} = 3.1$ and 3.8%, respectively).

Genetic Relationships among Populations

The genetic relationship among populations were studied by principal component analysis and multidimensional scaling. Integral parameters (factors) that determine the allele frequency variation were identi-

fied by principal component analysis. Positions of the ethnic groups studied in the space of two principal components are shown in Fig. 2a.

The first principal component was responsible for 17% of total variation and discriminated well between Slavic groups (Russians and Ukrainians) and all other populations; the second component reflected 1.7% of the total variations. The obtained plot contained three groups that clearly differ in the magnitude of the Caucasoid component. The first group included Russians and Ukrainians, the second one comprised Uralic populations, and the third one included Mongoloid ethnic groups.

To investigate the relationships between the populations of Siberia, Central Asia, and Far East in greater detail, four ethnic groups speaking Indo-European and Finno-Ugric languages (Russians, Ukrainians, Khanty, and Mansi) were excluded from analysis (Fig. 2b). In the resulting plot, two population groups were identi-

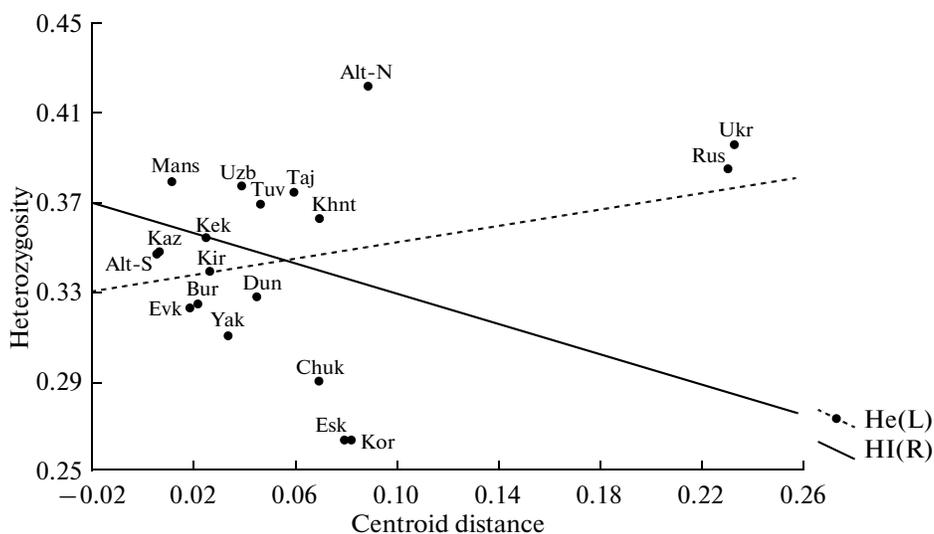


Fig. 3. Plot of heterozygosity vs. centroid distances of allele frequencies based on allele frequencies at eight autosomal polymorphic Alu loci. He(L) is fat dots and the dotted line show the observed values of expected heterozygosity and the corresponding regression line; HI(R) is solid line is the regression line for heterozygosity predicted from Harpending–Ward model.

fied. Of the two Caucasoid populations of Central Asia, Uzbeks clustered with Altaians, in agreement with linguistic classification: both these ethnic groups speak Turkic languages of the Altaic family. Within the other group, which included Mongoloid populations, several smaller clusters could be identified. In our chart, Evenks are positioned near Southern Altaians, Eskimo are close to Chukchi, and Tuvans are near Nivkh. Tajiks occupy an isolated position in the principal component space, similarly to Eastern European populations (Russians and Ukrainians) in Fig. 2a, which probably reflects an ancient Indo-Iranian component in their gene pool. The positioning of Kets and Koryaks is probably related to the presence of an ancient Paleosiberian component.

Analysis of Gene Flow

Ethnic groups included in this study differ in the intensity of genetic events (first of all, migrations) they have experienced, and the respective modern populations differ in levels of gene exchange with their neighbors. The relative intensity of gene flow in populations of northern Eurasia was evaluated using the approach developed by Harpending and Ward [31] based on Wright's island model [42].

Figure 3 shows the results obtained in our analysis of 19 ethnic groups. The solid line represents the theoretically expected dependence between the distance from the centroid and heterozygosity. In populations lying below this line, gene flows are weaker than theoretically expected, while in those above this line, they are stronger. Our results suggest that Russians, Ukrainians, Tajiks, Uzbeks, Tuvans, Northern Altaians, Khanty, and Mansi have been recipients of a consider-

able gene flow from external populations. Other populations are more isolated. The position of Russians in the plot reflects the history of the Russian population of Siberia formed by migrations from European Russia and the mixed character of populations included in the study. To explain the fact that Tuvan and Northern Altaian populations also lie above the expected heterozygosity line, it should be noted that these ethnic groups originated in the circumstances where large population groups repeatedly migrated across the Altai and Sayan region in different historical periods and in the prehistorical time. Peoples of eastern Siberia and the Far East are characterized with the lowest levels of gene exchange with other island populations, since gene pools of eastern Siberian ethnic groups have probably evolved in the conditions of strong isolation by distance and were little affected by migrations.

Correlations between Genetic Diversity and Geographic, Linguistic, and Anthropological Distances

Correlation analysis was performed using the Mantel's test, which evaluates the extent and significance of correlation between two matrices that represent genetic distances among populations or ethnic groups or any other variation that can be expressed as pairwise distances [32]. Matrices of genetic, anthropological, and linguistic distances among populations were constructed as described in [2].

We analyzed the correlation between the matrix of genetic distances based on Alu insertion polymorphisms and the matrices of geographic, linguistic, and anthropological distances (Table 5), and found that interpopulation divergence by autosomal markers was

Table 5. Matrix correlation and respective portions of genetic variation

Matrix	Correlation coefficient	Portion of genetic variation	Significance
Geographic distances	0.5422	0.2945	0.0000
Linguistic distances (classification)	0.3157	0.0990	0.001
Anthropological distances (classification)	0.6170	0.3807	0.0000
Anthropological distances (dermatoglyphics)	0.4047	0.1638	0.002

significantly correlated with differentiation by all other parameters. The highest levels of significance were observed for a comparison of the genetic distance matrix to the matrices of geographic distances and anthropological classification. The correlation between genetic and anthropological distances was the highest (0.617), while genetic and linguistic distances were correlated with the lowest coefficient (0.316). The portion of genetic variation associated with anthropological factors was also the highest (38%).

In general, our results indicate that genetic distances determined by Alu insertion polymorphisms have the strongest correlation with anthropological characteristics of populations in question. This is an expected finding, since, to a considerable extent, these characteristics are determined by genetics. In addition, the pattern of genetic differentiation is significantly affected by the territorial factor.

Thus, we have analyzed the gene pools of Northern Eurasian populations using a set of autosomal Alu repeat polymorphisms. Our results indicate that gene pools of populations from different regions differ considerably in genetic diversity, genetic differentiation, and degree of isolation. Populations of Eastern Europe, as well as southern and western Siberia, are characterized by high genetic diversity, strong migration effects, and a low level of genetic differentiation among populations. On the other hand, ethnic groups of eastern Siberia and the Far East show relatively low genetic diversity levels along with much stronger differentiation and little gene flow from external populations. A phylogenetic analysis suggested that the structure of the populations' gene pools changes gradually from west to east, and a correlation analysis indicated that anthropological and territorial factors have a significant impact on the pattern of genetic differences.

ACKNOWLEDGMENTS

This study was supported by the Federal Target Program "Human Resources in Science and Teaching in Innovative Russia" (agreement no. 8042) and the Russian Foundation for Basic Research (project no. 13-04-02023a).

REFERENCES

1. Limborskaya S.A., Khusnutdinova E.K., Balanovskaya E.V. 2002. *Etnogenomika i genogeografiya narodov Vostochnoi Evropy* (Ethnogenomics and Genogeography of the Peoples of Eastern Europe), Moscow: Nauka.
2. Stepanov V.A. 2002. *Etnogenomika naseleniya Severnoi Evrazii* (Ethnogenomics of the Population of Northern Eurasia), Tomsk: Pechatnaya Manufaktura.
3. Fedorova S.A. 2008. *Geneticheskie portrety narodov Respubliki Sakha (Yakutiya): analiz linii mitokondrial'noi DNK i Y-khromosomy*. (Genetic Portraits of the Peoples of the Republic of Sakha (Yakutia): Analysis of Mitochondrial DNA and Y-chromosome Lineages), Yakutsk: Yakutsk. Nauch. Tsent. Sib. Otd. Ross. Akad. Nauk.
4. Derenko M.V., Malyarchuk, B.A. 2010. *Molekulyarnaya filogeografiya naseleniya Severnoi Evrazii po dannym ob izmenchivosti mitokondrial'noi DNK* (Molecular Phylogeography of the Population of Northern Eurasia According to Data on Mitochondrial DNA Variation), Magadan: Severovost. Nauch. Tsent. Dal'nevost. Otd. Ross. Akad. Nauk.
5. Batzer M.A., Stoneking M., Alegria-Hartman H., et al. 1994. African origin of human-specific polymorphic Alu insertion. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12288–12292.
6. Hammer M.F., Karafet T.M., Redd A.J., et al. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203.
7. Wallace D.C. 1995. Mitochondrial DNA variation in human evolution, degenerative disease and aging. *Am. J. Hum. Genet.* **57**, 201–223.
8. Jorde L.B., Watkins W.S., Bamshad M.J., et al. 2000. The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988.
9. Jorde L.B., Wooding S. 2004. Genetic variation, classification and "race." *Nature Genet.* **36** (11), 28–33.
10. Zhivotovsky L.A., Rosenberg N.A., Feldman M.W., et al. 2003. Features of evolution and expansion of modern humans inferred from genome-wide microsatellite markers. *Am. J. Hum. Genet.* **72**, 1171–1186.
11. Tishkoff S.A., Kidd K.K. 2004. Implications of biogeography of human populations for "race" and medicine. *Nature Genet.* **36** (11), 21–27.
12. Cruciani F., Trombetta B., Massaia A., et al. 2011. A revised root for the human Y chromosome phylogenetic tree: The origin of patrilineal diversity in Africa. *Am. J. Hum. Genet.* **88** (6), 814–818.

13. Stoneking M., Fontius J.J., Clifford S.L., et al. 1997. Alu insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res.* **7**, 1061–1071.
14. Novick G.E., Novick C.C., Yunis J., et al. 1998. Polymorphic Alu insertions and the Asian origin of native American populations. *Hum. Biol.* **70** (1), 23–39.
15. Stepanov V.A., Puzyrev V.P., Spiridonova M.G., Khitrinskaya I.Yu. 1999. Analysis of the Alu insertion polymorphism in urban and rural Russian populations of Siberia. *Russ. J. Genet.* **35** (8), 979–983.
16. Khitrinskaya I.Yu., Stepanov V.A., Puzyrev V.P. 2001. Analysis of the Alu polymorphism in the Buryat populations. *Russ. J. Genet.* **37** (11), 1306–1311.
17. Khitrinskaya I.Yu., Stepanov V.A., Puzyrev V.P. 2003. Alu repeats in the human genome. *Mol. Biol. (Moscow)*. **37** (3), 325–333.
18. Khitrinskaya I.Yu., Stepanov V.A., Puzyrev V.P., Spiridonova M.G., Puzyrev K.V., Maksimova N.R., Nogovitsyna A.N. 2003. Genetic peculiarity of the Yakut population as inferred from autosomal loci. *Mol. Biol. (Moscow)*. **37** (2), 205–209.
19. Salimova A.Z., Kutuev I.A., Khusainova R.I., Akhmetova V.L., Svyatova G.S., Berezina G.M., Khusnutdinova E.K. 2005. Analysis of ethnogeographic groups of Kazakhs based on nuclear genome DNA polymorphism. *Russ. J. Genet.* **41** (7), 794–800.
20. Yunusbayev B., Kutuev I., Khusainova R., Guseinov G., Khusnutdinova E. 2006. Genetic structure of Dagestan populations: A study of 11 Alu insertion polymorphisms. *Hum. Biol.* **78** (4), 465–476.
21. Khusainova R.I., Akhmetova V.L., Kutuyev I.A., et al. 2004. Genetic structure of the Volga–Ural and Central Asian populations inferred from the data on Alu polymorphism. *Russ. J. Genet.* **40** (4), 443–449.
22. Solovieva D.S., Balanovska E.V., Kuznetsova M.A., Vasinskaya O.A., Frolova S.A., Pocheshkhova E.A., Evseeva I.V., Boldyreva M.N., Balanovsky O.P. 2010. The Russian gene pool: The gene geography of Alu insertions (ACE, APOA1, B65, PV92, TPA25). *Mol. Biol. (Moscow)*. **44** (3), 393–400.
23. Lavryashina M.B., Ul'yanova M.V., Balaganskaya O.A., et al. 2010. A genetic portrait of ten small nations of southern Siberia: 2. Gene pool heterozygosity and partitioning according to data on autosomal DNA markers. *Med. Genet.* **8** (9), 16–23.
24. Lavryashina M.B., Ul'yanova M.V., Vasinskaya O.A., et al. 2010. A genetic portrait of ten small nations of southern Siberia: 1. Polymorphism of autosomal DNA markers. *Med. Genet.* **8** (3), 12–17.
25. Lobov A.S., Kutuev I.A., Khidiyatova I.M., et al. 2008. Analysis of the genetic structure of Bashkir subpopulations based on polymorphic Alu insertion loci. *Med. Genet.* **7** (8), 30–37.
26. Bogotova Z.I., Kutuev I.A., Khusainova R.I., et al. 2010. Analysis of Alu insertion polymorphism in populations of Kabardians and Balkars. *Med. Genet.* **8** (1), 19–24.
27. Lependina I.N., Churnosov M.I., Atramentova L.A., Ischuk M.A., Tegako O.V., Tegako L.I., Balanovskaya E.V. 2008. The gene pool of Belgorod oblast population: Study of biochemical gene markers in populations of Ukraine and Belarus and the position of the Belgorod Population in the Eastern Slavic gene pool system. *Russ. J. Genet.* **44** (4), 484–494.
28. Lependina I.N., Balanovskaya E.V., Churnosov M.I. 2008. The gene pool of Belgorod oblast population: The distribution of immunological and biochemical gene markers. *Russ. J. Genet.* **44** (4), 470–483.
29. Johns M.B., Pauls-Thomas J.E. 1989. Purification of human genomic DNA from whole blood using sodium perchlorate in place of phenol. *Anal. Biochem.* **180**, 276–278.
30. Felsenstein J. 1993. *PHYLIP (Phylogeny Inference Package), Version 3.5c*. Seattle, Wash.: Univ. of Washington Press.
31. Harpending H.C., Ward R.H. 1982. Chemical systematics and human populations. In: *Biochemical Aspects of Evolutionary Biology*. Nitecki M.H. Ed. Chicago: Univ. of Chicago Press, pp. 213–256.
32. Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.
33. Nei M. 1987. *Molecular Evolutionary Genetics*. NY: Columbia Univ. Press.
34. Batzer M.A., Deininger P.L. 2002. Alu repeats and human genomic diversity. *Nature Rev. Genet.* **3**, 370–379.
35. Puzyrev V.P., Erdynieva L.S., Kucher A.N. 1999. *Genetiko-epidemiologicheskoe issledovanie naseleniya Tuva (Genetic–Epidemiological Examination of the Tuva Population)*, Tomsk: STT.
36. Alekseeva T.I. 1984. Specific anthropological features of recent Tuvinians: Cephalometry and cephaloscopy. In: *Antropoekologicheskie issledovaniya v Tuve (Anthropological Studies in Tuva)*. Alekseeva T.I., Uryson M.I. Eds. Moscow: Nauka, pp. 75–125.
37. Kucher A.N., Ondar E.A., Stepanov V.A., et al. 2003. *Tuvinty: Geny, demografiya, zdorov'e (Tuvinians: Genes, Demography, and Health)*, Tomsk: Pechatnaya Manufaktura.
38. Li J.Z., Absher D.M., Tang H., et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. **319**, 1100–1104.
39. Stepanov V.A., Prokhorchuk, E.B., Khusnutdinova, E.K., et al. 2010. Eurasian genetic landscape: Genogeography in the mirror of full-genome variability. *Materialy VI Rossiiskogo s'ezda obshchestva meditsinskikh genetikov (Proc. 6th Congr. Russ. Soc. Med. Genet.)*, Rostov-on-Don, p. 171.
40. *Genofond i genogeografiya narodonaseleniya (Gene Pool and Genogeography of Population)*, vol. 2: *Genogeograficheskii atlas naseleniya Rossii i sopredel'nykh stran (Genogeographic Atlas of the Populations of Russia and Bordering Countries)*. 2003. Rychkov Yu.G. Ed. St. Petersburg: Nauka.
41. Bruck S.I. 1986. *Naselenie mira. Etnodemograficheskiy spravochnik (The World Population Today: Ethnodemographic Processes)*. Moscow: Nauka.
42. Wright S. 1951. The genetical structure of populations. *Ann. Eugenics.* **15**, 323–354.

Translated by D. Timchenko