Overview of Genetic-epidemiological Studies in Ethnically and Demographically Diverse Isolates of Dagestan, Northern Caucasus, Russia

Kazima B. Bulayeva

Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

Aim To assess genetic diversity and genetic distances among isolated populations from Dagestan.

Methods A cross-population genetic epidemiology design was applied in ethnically and demographically diverse isolates from Dagestan, some with more than 200 and some with less than 100 generations of demographical history since their founding.

Results The analysis of genetic diversity showed that Dagestan ethnic populations are clearly close to European ethnic populations. The genetic data support the view of them as ancient, highly isolated populations 85%-97% the rate of the endogamy and inbreeding coefficient F = 0.010-0.015. Many Dagestan populations have very high prevalence of certain complex diseases such as cardiovascular illnesses, cancer, schizophrenia, mental retardation, and progressive muscular dystrophy. Lifetime morbid risk for schizophrenia in the isolates varied from 0 to 5%. Among the relatives, the number of men with chronic schizophrenia was at least twice as high as women. The average age of onset of schizophrenia was 21.2 years for offspring of consanguineous marriages and 17.4 years for offspring of non-consanguineous marriages ($P = 0.033$).

Conclusion The results support the hypothesis that cross-population design provides unique opportunities for observing reliable ancestral haplotypes with disease predisposing loci, as well as population-specific linked loci.

> **Correspondence to:**

Kazima Bulayeva
Vavilov Institute of General Genetics,
Russian Academy of Sciences
119991 Moscow, Russian Federation
bulaeva@vigg.ru

Mapping genes for complex diseases is one of the main priorities of modern genetics. This is because such diseases account for more than 90% of the global burden of disease and are responsible for a substantial proportion of premature deaths and disabilities. Genetic isolates provide an outstanding opportunity for identification of susceptibility genes for complex diseases. These isolates can be classified either as primary, with longer demographic history and constant total and effective sizes living in a stable environment, or as secondary, with a shorter demographic history and expanding size (1). Primary isolates still exist among ancient indigenous ethnic groups that are living in extreme and stable environments, such as deserts, highlands, jungles, or in the Arctic Circle. Secondary isolates are established by historically younger migrants (eg, Finnish isolates or Afrikaners) or religion sects (eg, Amish from North America or Mormons). It is known that ethnic diversity in human populations is related to their genetic diversity. Owing to these genetic characteristics of the primary isolates, the probability that linkage or linkage disequilibrium (LD) is caused by physical linkage between the marker and disease genes substantially increases in the primary isolates in comparison with secondary ones and especially in comparison with outbred, genetically heterogeneous populations.

Dagestan, which is located in the Northern Caucasus region of Russia, contains numerous primary and secondary isolates comprised of the 26 indigenous ethnic groups who have lived for hundreds of generations in the same stable highland region. Such ethnically diverse and demographically old genetic isolates could prove very important for the reliable detection of susceptibility genes, due to differences between isolates in ancestral haplotypes with pathogenic loci. Genes of minor or moderate effects should be more easily detectable in such ethnically homogeneous isolated populations as those in Dagestan, which have been isolated for long periods of time (2-4). Because the differences in the demographic age

of genetic isolates are related to the number of meioses and recombination events over generations, we hypothesized that primary isolates may have a relatively homogenous set of fewer disease-predisposing loci derived from a limited number of ancestors, which could make detection of disease susceptibility genes less time-consuming and more cost-effective.

Materials and methods

The Caucasus Mountains are situated astride of what must have been a major corridor of movement since the expansion of modern humans. Therefore, it may have functioned as a refuge and *cul-de-sac* off these migration streams (5). These ethnic groups belong to the four linguistic families – Dagestan-Nakh, South Caucasian, Indo-European, and Altaic (6). The southern two-thirds of Dagestan are located in the Main Caucasus Mountains, reaching 2000-4000 m above the sea level. The northern third is a flat plain that extends along the western shores of the Caspian Sea. The Republic of roughly 50 000 km² has a population of about two million people. While some of them are urban inhabitants, there are many isolated ethnic groups that rely on agriculture, herding, and craft production, especially in the almost inaccessible Caucasus Mountains part of Dagestan (5,6).

Observed ethnic populations are shown on the map of Dagestan (Figure 1). These ethnicities live in isolated villages that can be classified as primary isolates according to Neel's definition (1), because in each of these villages people have lived in the same highland environment for hundreds of generations. Their total population size has remained stable, and endogamy and inbreeding have occurred (7,8). Population size of individual isolates varies from 600 to 3000 (5). The smallest ethnic groups are often limited to a single unilingual village. In many ways, the geographically isolated mountain regions divided by natural barriers (or "auls", as they are referred to in this part
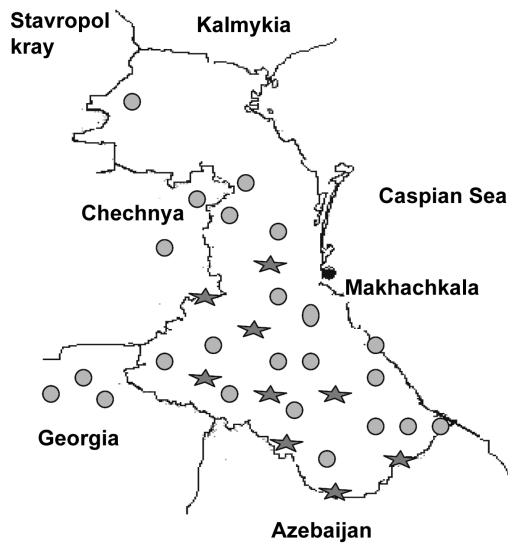
Figure 1. Map of Dagestan and surrounding regions. Studied ethnic populations (circles); isolated populations with aggregation of certain complex diseases (stars).

of Asia), represent independent experiments in ethnic differentiation: they have unique languages, traditions, arts, and customs (5,6). Examples of the mountain auls are Kubachi, the aul of goldsmiths; Gocatl, the silversmiths; Untsukul, the woodcarvers; Balhar, potters; Tsovkra, tightrope walkers; and the Kurah, the aul of unique carpet makers. There are also auls of tinsmiths, boot-makers, dancers, and singers. The main occupations among highlanders have been agriculture (primarily on hillside terraces) and stock raising (primarily sheep). Despite the harsh environmental conditions and limited terrain, the highlander groups have successfully farmed over many centuries (5,6). In fact, some of them may have contributed to the successful exploitation of the world's founder crops on the hillside terraces.

In our expedition studies, written informed consent was obtained from each participant who agreed to voluntary participation in our study and provided a blood sample and demographical data about types of marriages and reproductive parameters (9).

The set of short tandem repeats (STR) described in Jorde et al (4) in 1997 were typed in

the laboratory at the University of Utah, USA (4,10). Data collection, laboratory procedures and data analysis were described in great detail elsewhere (5,6,9,10). Most of the STR loci genotyped in the populations are tetra-nucleotides, but two are di- and one is a tri-nucleotide (10). We used these markers to assess genetic diversity and genetic distances among populations, and also to describe genetic diversity within populations, applying standard methods that were described in detail in earlier papers (6,10,11).

The isolates with aggregation of certain clinical phenotypes were ascertained based on state medical records in 1994-2004 expeditions. All mentally ill patients were diagnosed during hospitalization in two Dagestan psychiatric hospitals (12). Other complex diseases, such as cardiovascular diseases, cancer, and type 2 diabetes mellitus were diagnosed in central state and regional clinical hospitals. During our own expedition study, all clinical diagnoses of affected pedigree members were re-examined by the two experienced Dagestan specialists who participated in the expeditions, using a Russian translation of the structured psychiatric interview (12). Diagnostic Interview for Genetic Studies (DIGS), based on DSM-IV criteria, was used (12). The obtained diagnoses were compared and discrepancies were resolved by discussion or additional information was collected from a subject (12,13). Non-resolved cases were coded as unknown.

The prevalence for schizophrenia in the isolates was measured by the epidemiological index of lifetime morbid risk (LMR) which permits an estimation of the population "schizophrenia load" and allows for comparison between populations (14). The LMR coefficient is a probability that a subject will develop the disease until a certain age, which is 14-45 years for schizophrenia. This method includes a correction for sample sizes and takes into account the amount of unaffected subjects using the formula:

$$LMR = a/[N-(n_o + 0.5n_w)],$$

where a is the number of schizophrenic patients examined, N is the total population size, n_o is the number of subjects who are younger than the lower age limit for schizophrenia onset (children below 15 years of age), and n_w is the number of unaffected subjects whose current ages fall within the interval of schizophrenia onset (15-45 years) (14).

Results

Cluster analysis of genetic distances using 21 short tandem repeat polymorphisms (STRP) showed that the Dagestan ethnic groups were even closer to European and less like African and Asian populations than some European ethnic populations (Figure 2). The results of STRP loci heterozygosity analysis showed that all of Dagestan's ethnic populations are prominent outliers in homozygosity in terms of genetic distance from the African Centroid (Figure 3).
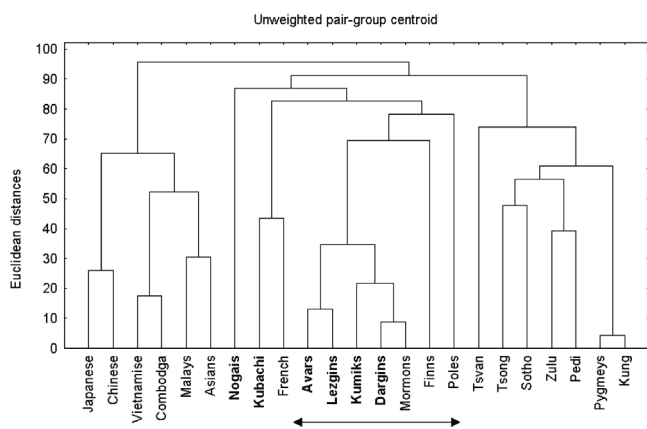

Figure 2. Cluster analysis of genetic distances between Dagestan and other global ethnic populations (using 21 short tandem repeat loci). Group of Mormons includes the subjects of Northern European origin. Group of Asians includes the subjects of Chinese, Japanese, Vietnamese, Cambodians, Koreans, Malayan, and Taiwanese ethnic origin.

The coefficient of inbreeding ranged from 0.0017 in the Kumik population to 0.0133 in the isolate of the smallest ethnicity, Tindals and Kubachians (Table 1), which are very high values at the level of entire populations.
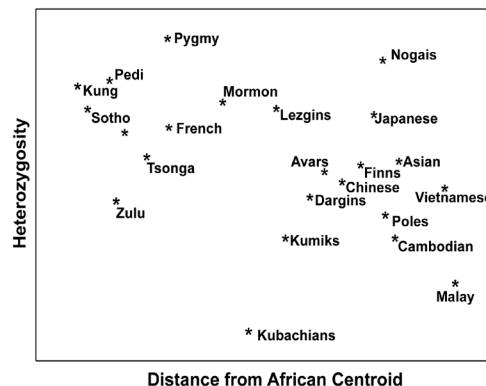

Figure 3. Heterozygosity level of the Dagestan and other world populations plotted against the genetic distance from the African Centroid.

Table 1. Rate of consanguineous marriages and inbreeding coefficient observed in Dagestan and in other ethnic groups in the world*

Ethnicity	Number of marriages	Proportion of consanguineous marriages	Average inbreeding coefficient
Dagestan:			
Botlikhs	101	16.7	0.0050
Kumiks	150	20.0	0.0017
Laks	223	47.3	0.0110
Avars	244	42.9	0.0103
Andians	166	31.6	0.01195
Tindals	170	32.9	0.01344
Dargins	350	23.7	0.0119
Lezgins	175	29.7	0.00991
Kubachians	98	26.4	0.0121
Tindals	158	31.7	0.0133
Siberia:			
Northern Khants	335	–	0.00058
Forest Nens	1600	–	0.012
Nganasans	700	–	0.002
Europe:			
French	530 000	0.7	0.00023
Italians	1 646 612	1.9	0.0007
Spain	2069	10.6	0.00275
North America:			
Canada-Catholics	51 729	1.5	0.00045
US Catholics	133 228	0.1	0.00008
US Mormons	625	9.9	0.00189
South America:			
Brazilians	212 090	4.8	0.00225
Costa-Ricans	3833	3.4	0.00114
Ecuadorians	3954	6.3	0.00229
Columbians	34 470	3.0	0.00119
Mexicans	28 292	1.3	0.00031
Peruvians	565	4.1	0.00279
Asia:			
Japanese	213 148	8.2	0.004
Indians	3520	11.4	0.00612
Israeli	11 424	9.7	0.00387
Africa:			
Egyptians	1782	75.8	0.03335
Guineans	1280	29.9	0.00819

*Source: review of the literature in public domain performed in reference 15.

Table 2 presents demographic and epidemiological description of some Dagestan highland ethnic isolates, most of which can be described as primary isolates. Those most demographically oldest isolates have a predominant aggregation of certain complex diseases such as schizophrenia, depression, cardiovascular diseases, and others (Table 2).

Table 3 describes the isolates with aggregation of the mental diseases. The obtained results show the lifetime morbid risk (LMR) of schizophrenia in the isolates where aggregation of schizophrenia varied from 0.0143 to 0.0495 (Table 3). An important clinical finding was that certain Dagestan isolates tend to have certain types of schizophrenia. For instance, the isolate #6022 was characterized by aggregation of the disorganized type of schizophrenia and a relatively early average age of onset (20.8 ± 1.51 years) with no cases of suicide. The study showed that most of the affected individuals in this isolate had grossly disorganized behavior, incoherence, marked loosening of association, flat or inappropriate affect, no sign of catatonic type of schizophrenia. The isolate #6007 was characterized by aggregation of paranoid schizophrenia and by a relatively later average age of onset (24.0 ± 2.35 years) and four suicides among affected individuals. The affected

Table 2. Epidemiology of chronic complex diseases in 11 different ethnic populations of Dagestan. For each isolate, a frequency of diagnosis of 11 types of complex chronic diseases among all encountered diagnoses of complex chronic diseases is presented*

Isolate code*	Ethnicity	Population size	Condition										
			schizophrenia	epilepsy	mental retardation	major recurrent depression	cardiovascular diseases	rheumatoid diseases	gastroenterological diseases	eye diseases	tuberculosis	cancer	neuromuscular dystrophy
6003	Dargins	1100	16.9	0.0	2.8	7.1	67.6	0.0	0.0	0.0	5.6	0.0	0.0
5017	Dargins	700	49.4	2.5	27.9	2.5	8.9	0.0	0.0	3.8	3.8	0.0	1.3
6002	Dargins	1190	5.2	0.0	2.6	5.2	74.3	0.0	0.0	2.6	7.5	0.0	2.6
3506	Dargins	1340	45.3	0.0	2.7	5.3	28.0	0.0	8.0	2.7	0.0	4.0	4.0
4034	Tindals	900	54.0	2.6	0.0	0.0	29.0	0.0	0.0	14.5	0.0	0.0	0.0
6009	Tindals	900	9.2	11.1	5.6	0.0	0.0	0.0	0.0	0.0	0.0	68.5	5.6
6005	Laks	818	5.7	0.0	8.0	52.3	4.6	0.0	0.0	13.6	15.9	0.0	0.0
6001	Avars	453	0.0	1.1	0.0	0.0	66.3	10.5	15.8	0.0	6.3	0.0	0.0
6011	Avars	750	0.0	2.5	2.5	0.0	12.5	15.0	0.0	0.0	7.5	0.0	60.0
3015	Kumiks	2000	55.6	2.7	0.0	0.0	19.5	0.0	0.0	0.0	11.1	11.1	0.0
3016	Lezgins	1450	7.5	0.0	5.0	0.0	71.5	12.0	4.0	0.0	0.0	0.0	0.0

*Codes given instead of the names of the villages for ethical reasons.

Table 3. Description of Dagestan genetic isolates with aggregation of different psychiatric diseases and conditions

Isolate code*	Ethnicity	Parameter								
		isolation time†	coefficient of inbreeding‡	population size	pedigree size§	NAF (NRC)‖	NRC¶	lifetime morbidity risk	mean age of onset of schizophrenia	condition
6005†	Laks	5000	0.0114	931	274	25 (19)	40	0.0241	22.1 ± 1.95	schizophrenia
6004	Laks	5000	0.0115	1125	375	15 (12)	28	0.0241	24.3 ± 2.20	unipolar depression
6007	Tindals	3000	0.0133	1800	539	36 (26)	86	0.0176	24.0 ± 2.35	schizophrenia
6029	Kumiks	600	0.0017	2000	342	24 (18)	43	0.0163	20.7 ± 1.61	schizophrenia
6010	Dargins	700	0.0066	2000	533	70 (24)	50	0.0297	20.6 ± 1.95	schizophrenia
6022	Dargins	4000	0.0091	1340	314	20 (16)	46	0.0271	20.8 ± 1.51	schizophrenia
6030	Dargins	3000	0.0110	700	257	28 (9)	14	0.0495	24.5 ± 2.14	schizophrenia
6015	Avars	800	0.0099	1153	396	23 (21)	43	0.0143	21.7 ± 2.05	schizophrenia
6006	Avars	3000	0.0112	1100	340	27 (16)	42	0.0310	21.1 ± 1.95	bipolar depression
6011	Dargins	700	0.0066	2000	187	15 (11)	26	–	–	mental retardation

*Codes given instead of the names of the villages for ethical reasons.
†Approximate isolation time of the populations known from archives, ancient publications and archeological data.
‡Estimated coefficient of inbreeding.
§number of persons within a larger pedigree.
‖ NAF (NRC) – number of affected individuals in the pedigree (and number of affected individuals that were recruited in the study).
¶NRC – number of individuals from the pedigree recruited in the study for clinical (DIGS) and DNA sampling, regardless of their affection status.

individuals in this isolate mostly had frequent auditory and sometimes visual hallucinations usually related to a single theme, preoccupation with one or several bizarre delusions, fears, and significant depressive episodes during affectations.

The minimum and maximum age of onset for schizophrenia across all the isolates varied from 13 to 40 years (mean age = 20.84±0.57). Mean age of onset of schizophrenia varied in different isolates from 20.6 ± 1.95 (isolate #6010) to 24.5 ± 2.14 (isolate #6030) (Table 3).

The reconstructed pedigrees comprised 317 to 539 persons from 9 to 12 generations. Our clinical study has found that only 139 out of 211 mentally ill patients, diagnosed with schizophrenia in Dagestan psychiatric hospitals have met the DSM-IV criteria of schizophrenia. The remaining 72 patients had, according to these criteria, various schizoaffective and affective disorders (eg, age-related affective disorders, anxiety neuroses, unipolar or bipolar psychoses, etc.).

Figure 4 shows a fragment of a pedigree with schizophrenia aggregation from the isolate with lifetime morbid risk of 4.95% (Table 1). Analysis of the marital structure in this and other pedigrees with aggregation of schizophrenia demonstrated that more than 60% of the affected subjects were offspring of consanguineous marriages. Testing the offspring of different types of marriages for variance homogeneity using the Levene test showed that the proportion

Figure 4. Sections of Dagestan primary isolate pedigree with aggregation of schizophrenia. Definitely affected subjects are colored, and probably affected subjects are marked by dots. Double lines indicate consanguineous marriages.

of schizophrenia patients was significantly higher among the offspring of consanguineous marriages (F = 9.55, d.f. = 1, P = 0.003). Analysis of the marital structure in some pedigrees with aggregation of schizophrenia demonstrated that more than 60% of the affected subjects were offspring of consanguineous marriages. Testing the offspring of different types of marriages for the variance homogeneity, using the Levene test, showed that the proportion of schizophrenic patients was significantly higher among the offspring of consanguineous marriages (F = 9.55, d.f. = 1, P = 0.003).

Discussion

The analysis of genetic distance that encompassed the Dagestan populations and groups belonging to the three major racial groups showed three distinct clusters – African, Asian, and European (including Dagestan) populations. Allele size variance in Dagestan populations was lower than that of other European populations. These two findings imply that these ethnic populations have indeed been relatively isolated and subject to more genetic drift than many other European populations (5).

In our previous study, it has been shown that a constant total and effective population size in the indigenous small ethnic groups is connected to selective pressures of extreme highland environment with hypoxia, high levels of radiation, abrupt climatic changes, and severe living conditions (eg, malnutrition and lack of medical care) (5).

Mutations responsible for complex diseases still persist in these isolates' current population because they have a significantly lower influence on the human viability, and therefore they are less affected by natural selection in comparison with Mendelian diseases (4). At the same time, due to cultural traditions of early marriages in Dagestan ethnic groups, an individual usually marries and already has several children before he
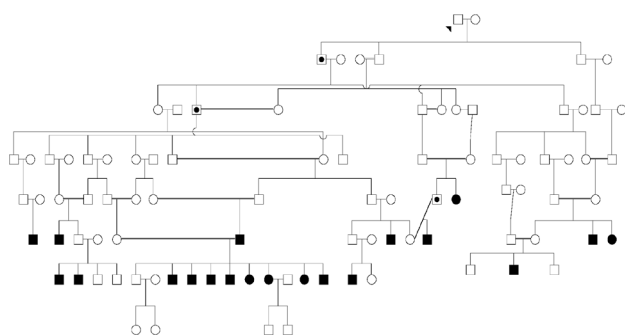
reaches the age of onset, and may also have more children during subsequent periods of affectation, since consanguineous marriages in the isolates are very seldom divorced. These traditions, along with founder effects and gene drift, have created a favorable environment for the enrichment of genes predisposing to complex diseases in the isolates. The consequence is apparent aggregation of certain complex diseases within the specific isolates. Our epidemiological study demonstrated that many of the isolates are characterized by the prevalence of certain complex diseases such as schizophrenia, bipolar and unipolar depression, mental retardation, cardiovascular diseases, and certain types of cancer. Furthermore, within certain diseases (such as schizophrenia), there was an apparent clinical heterogeneity of certain subtypes (5,12,13).

An important result of our study was the finding that, in the studied isolates, almost all patients with a homogeneous clinical phenotype belonged to one large pedigree with a limited number of founders. Obviously, these founders were also related, although this is difficult to ascertain. This aggregation of homogeneous clinical phenotypes in individual isolates may be explained in terms of the founder effect and endogamy/inbreeding.

In most cases, the symptoms of schizophrenia found during clinical examination in Dagestan psychiatric hospitals met the DSM-IV criteria that are widely accepted. However, we found that schizophrenia is over-diagnosed in these hospitals, whereas affective and schizoaffective disorders remain almost undiagnosed. Interestingly, the wide range of television-related delusions (0-40%) proved to be related to the social environment of the affected subjects in the studied isolates. This form of delusion was absent in those patients who had no TV sets at home because of poverty or religious reasons.

The population-specific aggregation of certain clinical phenotypes in genetic isolates, as well as of certain types of schizophrenia found in our study, may be explained in terms of the founder effect and genetic drift that favored significant aggregation of specific haplotypes containing disease-predisposing loci in some isolates and not in others. A given clinical phenotype that segregates with certain markers in some populations may segregate with other markers in other populations. The results obtained in cross-isolates differences in aggregation of certain subtypes of schizophrenia are consistent with earlier results reported by Gindilis et al (14). In their epidemiologic study, inter-population differences in prevalence of certain schizophrenia subtypes were found in an ethnic Komi population, a Russian secondary (religious) isolate, and in a mixed population of migrants.

Thus, in Daghestan primary isolates we have an opportunity for studying the causes of observed inter-population differences in epidemiology of various clinically apparent disease phenotypes related to founder effect, genetic drift, and strong isolation. In addition, this cross-population approach opens perspectives for detailed study of the relationship between the genetic (allele and locus) and clinical heterogeneity of complex diseases, which may promote the identification of their susceptibility genes.

References

1   Neel JV. Minority populations as genetic isolates: the interpretation of inbreeding results. In: Bittles AH, Roberts DF, editors. Minority populations: genetics, demography and health. London: Macmillan; 1992. p. 1-13.

2   Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, et al. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell. 1994;78:1073-87. Medline:7923357

3   Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996;58:1347-63. Medline:8651312

4   Jorde LB. Linkage disequilibrium and the search for complex disease genes. Genome Res. 2000;10:1435-44. Medline:11042143

5   Bulaeva KB, Pavlova TA, Kurbanov RM, Leal S, Bulaev OA. Genetic and epidemiological studies in Dagestan highland isolates [in Russian]. Genetika. 2003;39:413-22. Medline:12722642

6   Bulayeva K, Jorde LB, Ostler C, Watkins S, Bulayev O, Harpending H. Genetics and population history of Caucasus populations. Hum Biol. 2003;75:837-53. Medline:15018034

7   Bulayeva KB. Genetic basis of human psychophysiology. Moscow: Nauka; 1991.

8   Dubinin NP, Bulayeva KB. The study of genetic basis of the human individuality in different populations [in Russian]. Dokladi Akademii Nauk SSSR. 1982;265:334-7.

9   Bulayeva KB, Pavlova TA, Dubinin NP, Hay DA, Foley D. Phenotypic and genetic affinities among ethnic populations in Daghestan (Caucasus, Russia): a comparison of polymorphic, physical, neurophysiological and psychological traits. Ann Hum Biol. 1993;20:455-67. Medline:8215229

10 Bulaeva KB, Jorde L, Ostler C, Bulaev OA, Pavlova TA, Harpending H. STR polymorphism in populations of indigenous Daghestan ethnic groups [in Russian]. Genetika. 2004;40:691-703. Medline:15272568

11 Bulayeva K, Jorde L, Ostler C, Pavlova TA, Bulayev O, Harpending H. STR polymorphism in Caucasus ethnic populations. Russ J Genet. 2004;41:1-13.

12 Bulayeva KB, Leal SM, Pavlova TA, Kurbanov R, Coover S, Bulayev O, et al. The ascertainment of multiplex schizophrenia pedigrees from Daghestan genetic isolates (Northern Caucasus, Russia). Psychiatr Genet. 2000;10:67-72. Medline:10994643

13 Bulayeva KB, Leal SM, Pavlova TA, Kurbanov RM, Glatt SJ, Bulayev OA, et al. Mapping genes of complex psychiatric diseases in Daghestan genetic isolates. Am J Med Genet B Neuropsychiatr Genet. 2005;132:76-84. Medline:15389762

14 Gindilis VM, Gainullin RG, Shmaonova LM. Genetic and demographic patterns of the prevalence of various forms of endogenous psychoses. Russ J Genet. 1989;25:734-42.

15 Gadjiev A. Anthropology of small Daghestan populations [in Russian]. Makhachkala: Daghestan Branch of the USSR Academy of Sciences; 1971.